

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:40:42 ; Search time 31.43 Seconds
(without alignments)
1678.653 Million cell updates/sec

Title: US-09-701-572-2
Perfect score: 255
Sequence: 1 MDGTGNRPPTSTVRDNP.....KSTSEIGALSIGRTTIR 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq.032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2555	100.0	475	AAV53985
2	1287.5	50.4	496	AAAB43249
3	1276	49.9	478	ABBS59563
4	1231.5	48.2	1313	ABG13562
5	995	38.9	451	ABR6028
6	917	35.9	526	ABR60522
7	900	35.2	499	AAAR78656
8	893.5	35.0	499	AAAR78657
9	888	34.8	511	AAAB43592
10	835	32.7	855	AAAG39932
11	835	32.7	894	AAAG39931

12	835	32.7	898	21	AAAG39930	Arabidopsis thalia
13	254	9.9	74	22	ABG13561	Novel human diagno
14	233	9.1	478	21	AAAG35735	Zea mays protein f
15	229	9.0	479	21	AAAY79678	Drosophila Notchle
16	229	9.0	481	22	ABBS59486	Drosophila melanog
17	228.5	8.9	514	16	AAAR85881	WD-40 domain-contg
18	227.5	8.9	485	22	AAAB68284	Amino acid sequenc
19	226.5	8.9	485	22	AAAB92844	Human protein sequ
20	226.5	8.9	485	22	AAAB68516	Human GTP-binding
21	225.5	8.8	490	21	AAAG31339	Arabidopsis thalia
22	225.5	8.8	490	21	AAAG6519	Arabidopsis thalia
23	222.5	8.7	484	22	AAAB68282	Amino acid sequenc
24	221.5	8.7	439	22	ABBS7838	Drosophila melanog
25	218.5	8.6	471	21	AAAG23141	Arabidopsis thalia
26	218.5	8.6	471	21	AAAG6521	Arabidopsis thalia
27	218.5	8.6	473	21	AAAG3140	Arabidopsis thalia
28	218.5	8.6	473	21	AAAG6520	Arabidopsis thalia
29	215	8.4	514	20	AAAR84268	Candida albicans T
30	212.5	8.3	949	22	ABBS3108	Drosophila melanog
31	209	8.2	486	21	AAAG2985	Arabidopsis thalia
32	209	8.2	499	21	AAAG2984	Arabidopsis thalia
33	209	8.2	700	22	ABBS60376	Drosophila melanog
34	208	8.1	1326	22	ABBS7237	Drosophila melanog
35	208	8.1	1326	22	ABBS7238	Drosophila melanog
36	208	8.1	1326	22	ABBS7051	Drosophila melanog
37	206.5	8.1	339	19	AAAS5888	Human lipid metabo
38	206.5	8.1	339	19	AAAS5957	Human clao-1 prote
39	206	8.1	486	21	AAAG14448	Arabidopsis thalia
40	206	8.1	499	21	AAAG14447	Arabidopsis thalia
41	202.5	7.9	514	22	AAAB5225	Human protein sequ
42	196.5	7.7	261	22	ABBS68576	Drosophila melanog
43	196	7.7	361	22	ABBS68283	Amino acid sequenc
44	196	7.7	680	22	ABBS4206	Drosophila melanog
45	195.5	7.7	347	22	ABBS5902	Drosophila melanog

ALIGNMENTS

RESULT 1	AAV53985	standard; Protein: 475 AA.
ID	AAV53985	
AC	AAV53985	
XX		
DT	13-MAR-2000	(first entry)
XX		
DE	Amino acid sequence of an alfalfa fcr protein designated ccc52ms.	
XX		
KW	Alfalfa: ccc52ms; WD40 motif; flizy-related; fcr subfamily; fcr protein; cellular differentiation; endoreplication; cell proliferation;	
KW	plant cell; somatic embryogenesis; endoploidy; reserve tissue;	
KW	in vitro plant regeneration.	
KW		
XX		
OS	Medicago sativa.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	Modified-site	43..46
FT	Modified-site	/note= "potential cyclin dependent kinase phosphorylation site"
FT	Modified-site	128..131
FT	Modified-site	/note= "potential cyclin dependent kinase phosphorylation site"
FT	Modified-site	144..147
FT	Modified-site	/note= "potential cyclin dependent kinase phosphorylation site"
XX		
PN	FR2779433-AL.	
XX		
XX	10-DEC-1999.	
PD		
XX		
XX	08-JUN-1998;	98FR-0007174.
XX		

PR 08-JUN-1998; 98FR-0007174.
 XX
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Kondorosi E, Cebolla A, Kondorosi A;
 PI WPI: 2000-089743/08.
 DR N-PSDB; AA237024.
 XX
 XX New plant protein, useful for regulating differentiation and
 PT proliferation of plant cells, especially for stimulation of somatic
 PS embryogenesis
 XX
 XX Claim 2; Page 24-25; 36pp; French.
 CC The present sequence represents an alfalfa protein, designated ccc52ms.
 CC The protein contains repeated WD40 (TTP-Asp) motifs, and belongs to
 CC the fzr (fizzy-related) subfamily. The WD40 motifs are situated in
 CC the central portions and the C-terminal portion of the protein. The
 CC protein comprises 7 domains, and the last domain, domain 7, contains
 CC a potential liaison site for cyclins. The N-terminal region of the
 CC protein has a peptide sequence (DRFIPSR) which corresponds to a motif
 CC which is conserved in fzr proteins. The protein intervenes in the
 CC regulation of cellular differentiation, and augments endoreplication.
 CC The polynucleotide sequence and its antisense sequences are used to
 CC regulate the differentiation and proliferation of plant cells and for
 CC stimulating somatic embryogenesis. They are also useful for promoting
 CC endopolydly in plant cells or tissues to increase productivity of
 CC reserve tissues, and to stimulate in vitro regeneration of plants from
 CC callus in culture, by increasing the conversion of embryos.
 CC
 XX Sequence 475 AA:
 SO
 Query Match 100.0%; Score 2555; DB 21; Length 475;
 Best local similarity 100.0%; Pred. No. 8.8e-231;
 Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 0;
 QY 1 MDGTGNRNPPTSTVRDSSPPPEPESLRLHYSRMINSNHYTSPTYSRDFTPRSAS 60
 DB 1 mdgtgnrnpptstvrdrsspppepessrlhrsmnsnhytsptysrdftrpsas 60
 QY 61 KFAFDINTPTREGDSSSAVTLRLTALFGBDVAGPVTEKTDSPSMTLPNRIERYKT 120
 DB 61 kfaldintptregdrsssaytllrltalfgpdvagvptektdspmtlpnrnlfrykt 120
 QY 121 ETRGSMHSLSPMDDDFPVGVNHSVPKARVPRSPYVLDAPALQDDEFYLVDMWSHN 180
 DB 121 etrgsmhslspmdddfpvgvnhspvkarvprspykvlldapalqddfylnldvsshn 180
 QY 181 VLAVALGNCVYLWNAACSSKVTKLDGLGVDVCVSGVMAQRGTHLAVGNMKKVOIWDAR 240
 DB 181 vlavalgncvylwnaacsskvtklcdlgvdvcvsgvmaqrghthlavgnmkkvqiwdar 240
 QY 241 CKKISMGBHRLVGCALWSSSLSSGGKDKNIYORDIRTOEDFVSKISGKSEVCGLKW 300
 DB 241 ckkismgbhrlvgcalwssslssggkdkniyordirtgedfyskiskgsksevcgklkw 300
 QY 301 SYDNEELASGGNDKLFVWNOHSTQPVLYKCEHRAAVALAMSPRLHGLASGGGTADRC 360
 DB 301 sydneelasgndklfvmnohstqpvlykcehraavalaawspmlhglasgggtadrc 360
 QY 361 IRFMMTTNSHLSCHDTSQVCNLVWSKNVELVSTHGYSONOITVWYPPMSKLTATJTG 420
 DB 361 irfmmttntshlschdtsqvcnlywsknvelvsthgysngqitlvwypmsklatltj 420
 QY 421 HTYRLVLYLAISPDGQITVYAGDETLRFWMVPPSPKSONTESELSGLGRTTIR 475
 DB 421 htyrlvlylaispdgqitvtagdetlrfwmvppspksgnteselsglgalslgtttr 475

RESULT 2
 AAB43249

ID AAB43249 standard; Protein; 496 AA.
 XX
 AC AAB43249;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3013 polypeptide sequence SEQ ID NO:6026.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparkinsonian; antiparkinsonian; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 XX 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 PI WPI: 2000-602362/57.
 DR N-PSDB; AAC77458.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 5209-5211; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparkinsonian; antiparkinsonian; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antithyroid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation, to inhibit thrombosis; and as a contraceptive.

Sequence 496 AA:

Query Match 50.4%; Score 1287.5; DB 21; Length 496;
 Best Local Similarity 54.2%; Pred. No. 7.2e-112;
 Matches 269; Conservative 64; Mismatches 110; Indels 53; Gaps 14;

6 NRNPSTVSDNPPPEPSPESLRHVSRLMINSNHTYSPRTIYSDRFRISRSASKEAL- 64
 17 nentmrvtrmlrtrlttspssvys-----spk--hgdrffipraganawsvn 60

65 -FDIN---TPTEGR--DDSS-----ATTLLRTALRG---PDVAGPYTPKTKTSPS 107
 61 fhrlneknspsqnrkakatsdngkdqldaysallknellyagiekvqdpqtdrrllqps.120

108 MTLPRNRNFRKTKETROSM-----HSLSPMDDFVPGVNHSPYKARVRSRSPYK 158
 121 -lpekkglfytyslckrtpddgndvpspsvsnks--qklirspkrtrkskipik 177

159 VLDAPALODDFEYLNLVDSSHNVLAVGLGNCVYLMNACSSKVTYKLCDLGVD-DCVCSVGM 217
 178 vldapeidqddfylnlvdswslnvlsvglctvylswactsgvtrlcdlsvgeqdsvtswgm 237

218 AORGHILAVGTRNNKGVQIWDARCKIRSMGHRRLRVGALAMSSSLSSGGGRDNITVORD 277
 238 sergnlvavgtlkgfvtwdaagkksmlqgharvagalaawaeqslssgsrdtmllqrd 297

278 IRT---QEDFVSKLSGHSSEVGLKWSYDNRELASGDNKLFVWNOHSTQPVLYKTCENT 334
 298 ittpdlqge--rtlqghdvcevgklwscdhqlasggnndkllvwnhsislrvpqytleh 355

335 AAVKAIAMSPHLGLLASSGGGTADRCIRFWNTTNSHLSGMDTGSQVCNLYMSKNVLELV 394
 356 aavkalawspbhqyllaasgggtadrcirfwntltqplqgicldtsgqvcnllawskhanelv 415

395 STHGSONOITVWRYPMTSKLATLFGHRYRYLYLAISPDGQITVYGADETLRFMNVPSP 454
 416 sthgyssqgqyllwkypsltqvakltqhsyryvlylamspdgealvtgagdelcrlfwnvtsk 475

455 PKSONTESE-IGALSL 469
 476 trstkvkwsesvlnl 491

RESULT 3
 ID ABB59563 standard; Protein; 478 AA.
 AC ABB59563;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 5481.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PMD, Myers EW;
 PI WPI: 2001-656860/75.
 DR N-PsDB; ABL03666.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 5481; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB573737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 478 AA:

Query Match 49.9%; Score 1276; DB 22; Length 478;
 Best Local Similarity 54.9%; Pred. No. 8.1e-111;
 Matches 260; Conservative 62; Mismatches 114; Indels 38; Gaps 11;

22 PEPSPELRHVSRLMINS--NHTYSPRTIYSDRFRISRS---ASKFALFDIN----- 68
 4 peyekrlkhysparnllfnfessltptsltdrlfporynnwqtnfa--slknsdnp 61

69 -TPTEGRD-----DSSAYTTLRTALFGPDV-----AGVTPTEKNDPSMTLPNNRIFR 117
 62 qtskkgtdcgetardslayscalellknellysaiddvktageeremnytpa---akrsifk 118

118 YKETROSM-----HSLSPMDDFVPGVNHSPYKARVRSRSPYKVLADAPALODDFEYLNL 172
 119 ygsptkqdygecepslspvaks--qklirspkrtrksisrlpfvldapeidqddfyln 176

173 LVDMSSHNVLAVGLGNCVYLMNACSSKVTYKLCDLGVD-DCVCSVGNARQTHILAVGNNG 231
 177 lvdwssqnvilavglgscvylswactsgvtrlcdlspdanlvtswmergnvavagvhng 236

232 KVOIWDARCKIRSMGHRRLRVGALAMSSSLSSGGGRDNITVORDIRT-QEDFVSKLSG 290
 237 yvtwvdaankqklnghaarvagalaawnsdillsgsrdtwlqdrtrpqlserrlag 296

291 HKSEVCGLKWSYDNRELASGDNKLFVWNOHSTQPVLYKTCERTAAVKAIAKSPHLHGLL 350
 297 hrgevcglkwspdhqylasggnndrlyvwnqhsvnpvqsytehmaavkalawspbhngll 356

351 ASGGGTADRCIRFWNTTNSHLSGMDTGSQVCNLYMSKNVLELVSTHGYSONOITVWRYP 410
 357 asgggtadrcirfwntltqgpmgcvdcsqvcnllawskhselvtngsqnqyllwkyk 416

411 TMSKIATLTGHTVRYLYLAISPDGQITVYGADETLRFMNVPSPKSONTESE 464
 417 sltqvaktlghsyrylylalspdgealvtgagdelcrlfwnvtskarsqkenkv 470

RESULT 4
 ID ABB13562 standard; Protein; 1313 AA.
 AC ABB13562;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #13553.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX

PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001MO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
P1	Drmanac RT, Liu C, Tang YF;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS77749.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	

PS Claim 20, SEQ ID No 43921; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AB600010-AB630377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1313 AA: *

[illegible][illegible]

RESULT 5
 ABB66028 ID ABB66028 standard; Protein; 451 AA.
 XX
 AC ABB66028;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 24876.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 MN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EM;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL10131.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure: SEQ ID NO 24876; 21pp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 451 AA:
 SQ

Query Match 38.9%; Score 995; DB 22; Length 451;

PR 14-FEB-1994; 94US-0195730.
 XX (AMGE-) AMGEN INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX
 PI Weinstein J;

DR WPI; 1995-293119/38.
 DR N-PSDB; AAQ96099.

PT Mammalian cell cycle protein p55CDC and DNA sequences encoding it
 PT and methods of modulating cell division using cpds, effecting p55CDC
 PT or associated complexes useful in e.g. chemotherapy
 XX
 XX

Claim 1; Page 54-56; 87pp; English.

CC The p55CDC polypeptide is involved in mammalian cell division. DNA,
 CC proteins and antibodies derived from the polypeptide can be used to
 CC modulate cell cycle activity e.g. in chemotherapy, to inhibit the
 CC growth of tumour cells.
 XX
 XX

SQ Sequence 499 AA;

Query Match 35.28; Score 900; DB 16; Length 499;
 Best Local Similarity 39.28; Pred. No. 1.6e-75;
 Matches 195; Conservative 78; Mismatches 175; Indels 50; Gaps 9;

QY 8 NPPSTVIRDNSSPPPEPSRLRVS-----MINSNHTSPRTIYSDRFI 54
 DB 22 naplarwqrakatepapepmraanshsagrtptgpknskvgtlpskp-ggeryl 80
 QY 55 PPSASAK-----FALFDINTPTGSRDDSSAYTTTLRTALFGPDVAGPVTPKTDSPSM 109
 DB 81 pprsaagmevasfllskengpedggtprkkehqkawaalnfgdv----- 125

QY 110 LPRNIFRYKTEETROS-----MHSLSPFMDDFVPGVNHSPVKARPKVRSPIYLDAPAL 165
 DB 126 -eeakllrlsgkpnapegygnrlkylvsqkatpg---ssrkacrylpslptdlidape 181

QY 166 QDDFYRLNLDWMSHNVLAVGLGNCVYLMNACSSKVTKLDL-GVDDCVCSVGAORGTIL 224
 DB 182 rldylnlvdwssgnvlavaldnsyvlwmsgddllqlmqegpdyllssvawikegnyl 241

QY 225 AVCTNGKVOIMDARCKIRSMEGHRLRVGALAMSSSLSSGGRDKNTYQDRTQEDF 284
 DB 242 avgtlsaeqglwdvgqgkrlrmtshsartvsslswnsyllssgsrghlhhdvraehh 301

QY 285 VSKLSGHKSEVCGKLKWSYDNRELASGDNKLEFV---NQHSTQPVLYKCYCENTAAVKA 340
 DB 302 valtsghsgevcglrwapdgrhlasgndhlnvwpspgpegswvplqftqhgavav 361

QY 341 AMSPHLHGLASGGGTADRCIRFWNTTNSHLSCHMTGSOVCNLYWNSKVNDELVSTHGS 400
 DB 362 acpmpgsnllatvggtsdhlrlmncsacslavdhsqvcslwshpkykellsgfhga 421

QY 401 QNOIIVWRPYTMSKLTATLTGHTYRVLYLAISPQGTIVTGAGDETLRFNWNF---PSK 457
 DB 422 gnglviwkyptmakveelkgharvylstlmspdgatvasaadellrlwrfeldpallr 481

QY 458 QNTESEIGALSIGRTTIR 475
 DB 482 erekastskssllhglt 499

RESULT 8
 AAR78657
 ID AAR78657 standard; Protein; 499 AA.

XX AAR78657;
 XX
 DT 23-JAN-1996 (first entry)
 XX

DE Human cell cycle protein p55CDC.

XX Cell cycle; protein; p55CDC; antibody; chemotherapy; modulation;
 KW inhibition; growth.
 XX
 XX

OS Homo sapiens.

XX WO9521917-A1.

PD 17-AUG-1995.

PF 13-FEB-1995; 95WO-US01806.

PR 14-FEB-1994; 94US-0195730.

XX (AMGE-) AMGEN INC.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX

PI Weinstein J;

DR WPI; 1995-293119/38.
 DR N-PSDB; AAQ96100.

PT Mammalian cell cycle protein p55CDC and DNA sequences encoding it
 PT and methods of modulating cell division using cpds, effecting p55CDC
 PT or associated complexes useful in e.g. chemotherapy

Claim 1; Page 58-60; 87pp; English.

CC The p55CDC polypeptide is involved in mammalian cell division. DNA,
 CC proteins and antibodies derived from the polypeptide can be used to
 CC modulate cell cycle activity e.g. in chemotherapy, to inhibit the
 CC growth of tumour cells.
 XX
 XX

SQ Sequence 499 AA;

Query Match 35.08; Score 893.5; DB 16; Length 499;
 Best Local Similarity 38.58; Pred. No. 6.6e-75;
 Matches 187; Conservative 81; Mismatches 165; Indels 53; Gaps 8;

QY 8 NPPSTVIRDNSSPPPEPSRLRVS-----MINSNHTSPRTIYSDRFI 54
 DB 22 naplarwqrakatepapepmraanshsagrtptgpknskvgtlpskp-ggeryl 80

QY 55 PPSASAKFALDINTPTGSRDDSSAYTTTLRTALFGPDVAGPVTPKTDSPSKTLPNRN 114
 DB 81 pprsaagmevasfllskengsemsq-----tpkkehqkawaaln 121

QY 115 -----IFRYKTEETROS-----MHSLSPFMDDFVPGVNHSPVKARPKVRSPIYLDAPAL 162
 DB 122 gfdveeakllrlsgkpnapegygnrlkylvsqkatpg---ssrkacrylpslptdlidape 178

QY 163 PALDDFYRLNLDWMSHNVLAVGLGNCVYLMNACSSKVTKLDL-GVDDCVCSVGAORGTIL 221
 DB 179 pelrldylnlvdwssgnvlavaldnsyvlwmsgddllqlmqegpdyllssvawikeg 238

QY 222 THLAVTNGKVOIMDARCKIRSMEGHRLRVGALAMSSSLSSGGRDKNTYQDRTQEDF 281
 DB 239 nylavgtssaeqglwdvgqgkrlrmtshsartvsslswnsyllssgsrghlhhdvraehh 298

QY 282 EDEVSKLSGHKSEVCGKLKWSYDNRELASGDNKLEFV---NQHSTQPVLYKCYCENTAAV 337
 DB 299 ehvvaltsghsgevcglrwapdgrhlasgndhlnvwpspgpegswvplqftqhgavav 358

QY 338 KATAMSPHLHGLASGGGTADRCIRFWNTTNSHLSCHMTGSOVCNLYWNSKVNDELVSTH 397
 DB 359 kavawcpgsnllatvggtsdhlrlmncsacslavdhsqvcslwshpkykellsgfhga 418

QY 398 GYSQNOIIVWRPYTMSKLTATLTGHTYRVLYLAISPQGTIVTGAGDETLRFNWNF---PSK 457
 DB 419 gfaqnglviwkyptmakveelkgharvylstlmspdgatvasaadellrlwrfeldpallr 478

OY 458 QNTSE 463
Db 479 rrrere 484

RESULT 9
AAB43592
ID AAB43592 standard; Protein: 511 AA.

AC AAB43592;

DE 08-FEB-2001 (first entry)

Human cancer associated protein sequence SEQ ID NO:1037.

Human; cancer associated gene; cancer antigen; detection; cancer; diagnosis; cytostatic; proliferative; vulnerrary; immunomodulator; antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral; antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiant; dermatological; neuroprotective; thrombolytic; coagulant; nocitropic; vasoactive; antipsoriatic; antiangiogenic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening.

Homo sapiens.

MO200055350-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05882.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-587533/55.

N-PSDB; AAC77801.

Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -

Claim 11; Page 1624-1626; 2352pp; English.

AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnerrary; immunomodulator; antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiinflammatory; antihypertensive; antiallergic; antibacterial; antiviral; dermatological; neuroprotective; cardiant; thrombolytic; coagulant; nocitropic; vasoactive; antipsoriatic; antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

Sequence 511 AA;

Query Match 34.8%; Score 888; DB 21; Length 511;
Best Local Similarity 38.2%; Pred. No. 2.2e-74;
Matches 189; Conservative 80; Mismatches 158; Indels 68; Gaps 10;

OY 8 NPEPTSTVR-----DNSPPEPSPESLIRHVSRLMINSH-----YTSPS 45
Db 31 hpsmhpilragsksgsgrrepapsp-----wraanrshaagrtpttpkssskqtlps 84
OY 46 RTYSDFRIFPSRSASKAFALDINTPTREGDDSSSAYTTLRLTALFGPDVAGPVTPEKTD 105
Db 85 kp-ggdrylphrsaagmevasflskengpensq-----tpkkeh 124
OY 106 PSMTLPNRN-----IRPKTEFROS-----MHSLSFPMDDFVPGVNHSPVKARRP 153
Db 125 qkwalnlnngfdveaakllrlsgkpgnapgygnrlkvlysqkarpv---ssrkcterylp 181
OY 154 RSPYKVDNALDDPYLNLNVDWSSHNVLAVGLGNCVYLMNCSVTKLCPL-GVDDCV 212
Db 182 slpdrillapeirndylnlvdwsgnvlavaldnsyvlwsasgdlqlqmegpgeyl 241
OY 213 CSYGMARQRTHLAVGTNGKVKQITDAARCKIRSMEGHRLVGCALAWSSSLSSGGGRDN 272
Db 242 ssyawlkegnylavgtssaeavqlwvqgkrlrmctshaarygslswnsyilssgsrsgh 301
OY 273 IYQDRIRTOEDFVSKLSGHSKSEVCGLKMSYDNRELASGGDNKLFVW-----NQHSTQPV 328
Db 302 lnhdvraehhvatllsgshgvcgylrwapdgrflasgndnlvvnvpsapgegyvrlp 361
OY 329 KYCEHTAAVKAIAWSPHLHGLASGGTADRCIRFNNVTTSNLSHSCMDPQSOVCNLWMSK 388
Db 362 tftcqhgaavkavacpwsqnvlatggtldrhlrlmrvsgaelavdshgyscllwap 421
OY 389 NVNELVSTHGYSONQIILWRYPTMSKLATLTGHTYVLYLAISPDGQITVGADETLRF 448
Db 422 hkelisgshgfaeqnqlvlykypcmakvaeikghtsvlslmspdgavtasaadclrl 481
OY 449 WNVFPEKSGNTSE 463
Db 482 wrcefldparrrere 496

RESULT 10
AAG39932
ID AAG39932 standard; Protein: 855 AA.
AC AAG39932;
XX
DE 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49478.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145276.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161930.
PR	28-OCT-1999;	99US-0161982.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

RESULT 12
AAG3930
ID AAG3930 standard; Protein: 898 AA.
XX AAG3930;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SFG ID NO: 49476.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EPI03405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0125148.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130444.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 05-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140035.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR	20-AUG-1999;	9905-0149722.
PR	20-AUG-1999;	9905-0149723.
PR	20-AUG-1999;	9905-0149929.
PR	23-AUG-1999;	9905-0149902.
PR	23-AUG-1999;	9905-0149930.
PR	25-AUG-1999;	9905-0150566.
PR	26-AUG-1999;	9905-0150884.
PR	27-AUG-1999;	9905-0151065.
PR	27-AUG-1999;	9905-0151066.
PR	27-AUG-1999;	9905-0151080.
PR	30-AUG-1999;	9905-0151303.
PR	31-AUG-1999;	9905-0151438.
PR	01-SEP-1999;	9905-0151930.
PR	07-SEP-1999;	9905-0152363.
PR	10-SEP-1999;	9905-0153070.
PR	13-SEP-1999;	9905-0153758.
PR	15-SEP-1999;	9905-0154018.
PR	16-SEP-1999;	9905-0154039.
PR	20-SEP-1999;	9905-0154779.
PR	22-SEP-1999;	9905-0155139.
PR	23-SEP-1999;	9905-0155486.
PR	24-SEP-1999;	9905-0155659.
PR	28-SEP-1999;	9905-0156458.
PR	29-SEP-1999;	9905-0156596.
PR	04-OCT-1999;	9905-0157117.
PR	05-OCT-1999;	9905-0157753.
PR	06-OCT-1999;	9905-0157865.
PR	07-OCT-1999;	9905-0158029.
PR	08-OCT-1999;	9905-0158232.
PR	12-OCT-1999;	9905-0158369.
PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	13-OCT-1999;	9905-0159295.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161992.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

	Query Match	32.7%	Score 835;	DB 21;	Length 898;
	Best Local Similarity	40.7%;	Pred. No. 4.8e-69;		
	Matches 171; Conservative	68;	Mismatches 89;	Indels 92;	Gaps 7.
OY	51 DRFTSRASAKR--LDPINTPTBERDDSSAYTLLRLALGCPVAGVVTPEKIDSPSM	108			
	: : : :				
Dd	292 drfiprsamrIdyhfalteerkkqdsat-----tsert-----	327			
OY	109 TLPNNINIRYKETFQSHSLSPFDDDFEPCVNHSPVKAPRKVRSPRYKIYLDARLQDD	168			
	: : :				
Dd	328 -----ldapdiavdd	336			
OY	169 FYLNVDMSHHNVLAVGLNCVCYYLMNAASSKTKIKCLDCLVD-DVCYSVGMAQRGHILANG	227			
	: : : : :				

```

Db 337 fymlldwgsanvialadhevlylwdastgststslvtldeekpvtssinnapdgrhnavg 336b
OY 228 TNNKVOIWDARCKKTRISMF-GHRLRYGALAWSSLLSSGGRDKNIYORDIRTOEDFVS 286b
Db 397 Innsevgqlwdsasnrqtrtltkgqrsqrsajslawmhllltcgmglllindvrlrpsive 456b
OY 287 KLSGKHSKSVCGJAKMSYDREKLASGDNKKLEVMNO-----HSNQPALXKCEHTAAKAI 340b
Db 457 tyrhtgtgevgqlkxsgsqgqlaasgndvvhvldrvsaasnstqwlhrlehtsavkal 516b
OY 341 AMSPHLLGLLASGGGTADRCIRFNNVTNTNSSLSCMDYGSQVCNLWMSKNVELVSTHGYS 400b
Db 517 awcpeqanllatggggggtrtlkfnthtgcaclnsvdcgysclslwsknerellshgft 576b
OY 401 ONOIIWVWYPPMMSKATLTLGHTVYVVLKLAISPOOGTIVTGAGDET-----LRFNWVP 453b
Db 577 gngqlkwlkyspmvmaelctgtrtsvlylmaagppdgcvtasaagdelsspndkalsfwlpp 636b

```

CC	XX	RESULT 13
CC	XX	ABG13561
CC	XX	ID ABG13561 standard; Protein; 74 AA.
CC	XX	AC
CC	XX	ABG13561;
CC	XX	DT 18-FEB-2002 (first entry)
CC	XX	DE Novel human diagnostic protein #13552.
CC	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
CC	XX	food supplement; medical imaging; diagnostic; genetic disorder.
CC	XX	Homo sapiens.
CC	XX	MO200175067-A2.
CC	XX	PD 11-OCT-2001.
CC	XX	PF 30-MAR-2001; 2001WO-US08631.
CC	XX	PR 31-MAR-2000; 2000US-0540217.
CC	XX	PR 23-AUG-2000; 2000US-0649167.
CC	XX	(HYSE-) HYSEQ INC.
CC	XX	Dmanac RT, Liu C, Tang YF;
CC	XX	WPI: 2001-639362/73.
CC	XX	DR N-PSDB; AAS77748.
CC	XX	PT New isolated polynucleotide and encoded polypeptides, useful in
CC	XX	PT diagnostics, forensics, gene mapping, identification of mutations
CC	XX	PT responsible for genetic disorders or other traits and to assess
CC	XX	PT biodiversity -
CC	XX	Claim 20; SEQ ID No 43920; 103bp; English.
CC	XX	The invention relates to isolated polynucleotide (I) and
CC	XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	XX	and gene mapping, and in recombinant production of (II). The
CC	XX	polynucleotides are also used in diagnostics as expressed sequence tags
CC	XX	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	XX	to restore normal activity of (II) or to treat disease states involving
CC	XX	(II). (II) is useful for generating antibodies against it, detecting or
CC	XX	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	XX	a food supplement. (II) and its binding partners are useful in medical
CC	XX	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	XX	disorders involving aberrant protein expression or biological activity.
CC	XX	The polypeptide and polynucleotide sequences have applications in
CC	XX	diagnostics, forensics, gene mapping, identification of mutations
CC	XX	responsible for genetic disorders or other traits to assess biodiversity
CC	XX	and to produce other types of data and products dependent on DNA and


```

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158366.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

```

```

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 9.1%; Score 233; DB 21; Length 478;
 Best Local Similarity 20.9%; Pred. No. 5,4e-13;
 Matches 113; Conservative 76; Mismatches 161; Indels 190; Gaps 22;

```

QY 13 STVRDSSPPPEPESPELRHVSRLMNSHYTSPRTIYSDRIPR----- 57
DB 2 ataaadpppe--pslkkis-----lrktsl--dlfaphafllfpdaesklrt 50
QY 58 ----SASKFALFDINTPREGRDSSSAVYTLRLTALGPDVAGVPEPEKTDSPMTLPNR 113
DB 51 gckvaeygavkdl-tpeggr-----ggqgkypappst--alaipg- 89
QY 114 NIFRYKTEKRSMSLSFMDDFVPCVNSPVKPKKVPKSPK-----VIDAPALQ 166
DB 90 ----tqtkdang-----egsnaiypapimlpkpeestipgkntslspgss 133
QY 167 DDFYLNLY-----DMSHNYLAVGLG-----NCVY 191
DB 134 drfsalmerlprwprpwhapwkyrvlsglhwrvlratfpanewlctgsadtllk 193
QY 192 LMNACSSKVTYLCGLGDDCVCSVGMARGTTHLAVGTNNGKVOIWDARCKKIRSMEGHR 251
DB 194 lwdlaagtl-kltlthleqirglavagrhylfsagddkqvkwdleqkvltlryghl 252
QY 252 LRVGALAW--SSLSGSGRDKNIYORDIRFOEDFVSKLSGHSKSEVGLKMSYDNRELAS 309
DB 253 sgvyclalhpdlilllgygrdsvervwdltka-hvsaletghntvcsvfarptqgvte 311
QY 310 GGNDNKLFVWNOHSTQPVLYKCEHTAAVKAIAWSP-----HLH 347
DB 312 gshdtllkfwdlvagrmtclthhksvramalhpkekafasasadnvkfnlpkgeflh 371
QY 348 -----GILASGCGRADRCIRRWNTTSHLSCHMTGSGQVCNLMWSKN 389
DB 372 nmIsqgkllnsamvnedgylatcgdnsg--lwfwd-----wk-- 407
QY 390 VNELVSTHGYSQNOIIVWRYPTMSKSLATLNGHTYRVLYLAISPDGOTIVGADETLRFW 449
DB 408 ----sgnhfgqdgltv-qpgslesaciyalsydv-----sgsrlvsceadktilmw 454

RESULT 15
AAV79678
ID AAV79678 standard; Protein; 479 AA.
XX
AC AAV79678;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila Notchless protein.
XX
KW Notchless; Nle gene; Notch; signalling; neurodegenerative disease;
KM cancer; diagnosis; cytostatic; neuroprotective; therapy.
XX
OS Drosophila melanogaster.
XX
FH Key Location/Qualifiers
FT Domain 27..105 /note="Nle domain"
FT Misc-difference 67 /note="sequence deduced from nucleotide sequence
FT FT has an additional Lys residue between
FT Lys-67 and Ser-68"
FT Misc-difference 282 /note="encoded by CAA"
PN WO200026364-A1.

```

Job time: 339 sec

XX 11-MAY-2000.
PD 03-NOV-1999; 99WO-1B01891.
XX 03-NOV-1998; 98GB-0024045.
XX (EURO-) EURO MOLECULAR BIOLOGY LAB.
PA Cohen S, Boumeester A, Royet J;
PI WPI: 2000-365613/31.
DR N-PSDB; AAA27739.
XX Novel Notchless protein and nucleic acids encoding them useful for
PT treating and preventing cancer and neurodegenerative diseases
XX
PS Claim 1; Page 44; 52pp; English.
XX
XX The present sequence is that of Notchless, a novel protein of
CC Drosophila. Notchless was identified in a screen for dominant
CC modifiers of a Notch mutant phenotype in the Drosophila wing. The
CC mutant dominantly suppressed the wing notching phenotype of
CC notchoid mutations, and the Notchless protein was shown to bind to
CC the cytoplasmic domain of Notch. Notchless modified Notch
CC signaling activity in a variety of Notch-dependent signalling
CC process in both Drosophila and Xenopus embryos. The Notchless
CC protein has a novel highly conserved N-terminal domain followed by
CC 9 WD40 repeats. Notchless, and nucleic acids encoding it, can
CC be used in methods for the diagnosis and therapy of certain diseases,
CC particularly cancer and neurodegenerative diseases (claimed). A
CC Notchless mutant in a sensitised Notch genetic background is used
CC in a claimed method for identifying Notch compounds capable of modifying
CC the levels of expression or activity of a Notch protein.
XX
SQ Sequence 479 AA;

Query Match 9.08; Score 229; DB 21; Length 479;
Best Local Similarly 23.48; Pred. No. 1.3e-12;
Matches 91; Conservative 50; Mismatches 128; Indels 120; Gaps 13;

QY 164 ALDDPFLNLCVDMSSHNVLAVGLGNCVYLWNA-----CSKYTKLGLGVDPCVCS 214
DB 68 sledldiasvd--tenvld-----yqgqavfkvrvtrctsmg-----ghaavvs 115
QY 215 VQMAORGTHTLAVGTNNKQVQIWDARCKITRSMGHRIRVGLAWS--SSLSSGGGRDN 272
DB 116 Infspdgahlasgsgdtlvrtwdintetphctcghkqwlcvswapdgkrlasgckags 175
QY 273 IYORIRTOEDFVSKLGSCKSEVCGLKW-----SYDNRELASGGNDNKLFWNQHSTQPV 327
DB 176 lllwbpelqgqktrpilsghkhhincilawephrdpectrlasasgddcriwdvklqgcl 235
QY 328 LKYCEHTAVALKAIAMSPHLHGLASGGGTADRCIRFWNTT----- 367
DB 236 mlaaghtnavtavgg--agliyle--sskdrtykmwraadgllcrlfsgnahvnnial 291
QY 368 -----INSHLS-----CMDTGSQVCNLV----- 385
DB 292 stdyvlltgrfhrpvkdrskshlsteelqesalkrygavcpd---eveslvscednltl 348
QY 386 --WSKVNELVSTHGYSON-----OIIVRYPTMSKLAITLTGHT 422
DB 349 ylvrmnqncvemtgmghnvdvdkyspdklaasafdkasvrlwrasdgqymalfgrhv 408
QY 423 YRVLYLAISPDOFTVITGAGDETLRFWNV 451
DB 409 gavytvawsadsrlivsgskdsrlkwsv 437

Search completed: May 21, 2002, 11:46:21

Wed May 22 11:09:42 2002

us-09-701-572-2.rag

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:44:12 ; Search time 20.97 Seconds
(without alignments)
2176.559 Million cell updates/sec

Title: US-09-701-572-2

Perfect score: 2555

Sequence: 1 MGCTGNRNPPTSTYRDNSP.....KSONTESEIGALSGRTTIR 475

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1839	72.0	444	2	T05122
2	1825	71.4	482	2	T09351
3	1096	42.9	702	2	T27730
4	1066	41.7	556	2	T37680
5	985	38.6	421	2	T40614
6	933.5	36.5	566	2	S64005
7	923	36.1	447	2	T05977
8	920.5	36.0	457	2	T05978
9	901.5	35.3	450	2	T14352
10	900	35.2	499	2	B56021
11	893.5	35.0	499	2	A56021
12	876	34.3	440	2	T01768
13	777.5	30.4	488	2	T41719
14	674	26.4	610	2	S64126
15	441.5	17.3	509	2	T41034
16	431	16.9	535	2	S62437
17	331.5	13.0	849	2	T43432
18	311	12.2	1526	2	AC2239
19	310	12.1	336	2	T27762
20	278	10.9	676	2	AH2195
21	273.5	10.7	1747	2	AC1842
22	273	10.7	1258	2	A12155
23	272.5	10.7	786	2	AG3375
24	272	10.6	1356	2	T18521
25	269.5	10.5	304	2	AG1837
26	267	10.5	677	2	AE1861
27	256	10.0	1227	2	AE1810
28	254.5	10.0	614	2	S58306
29	250	9.8	586	2	T38892

ALIGNMENTS

30	245	9.6	1189	2	A12493	WD-repeat protein
31	244.5	9.6	559	2	AB2202	hypothetical prote
32	244.5	9.6	1683	2	AF2071	WD-40 repeat prote
33	242.5	9.5	1711	2	AD1842	WD-40 repeat prote
34	238.5	9.3	265	2	AF1890	WD-repeat protein
35	237	9.3	515	2	S19487	hypothetical prote
36	234.5	9.2	1708	2	AE1866	WD-40 repeat prote
37	231	9.0	589	2	AG2400	WD-repeat protein
38	224.5	8.8	934	2	AG1889	trp-asp repeat con
39	221.5	8.7	502	2	T41148	hypothetical WD-re
40	221.5	8.7	777	2	T41075	hypothetical prote
41	221	8.6	473	2	T33805	probable u4/u6 sma
42	217.5	8.5	462	2	T50168	WD-repeat protein
43	215.5	8.4	1189	2	AH2154	beta transductin-II
44	213.5	8.4	1049	2	T42045	WD-repeat protein
45	211.5	8.3	598	2	AE2415	WD-repeat protein

RESULT 1

T05122
hypothetical protein F7H19.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 26-May-2000
C:Accession: T05122
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
submitted to the Protein Sequence Database, July 1998
A:Reference number: 215399
A:Accession: T05122
A:Molecule type: DNA
A:Residues: 1-444 <BEV>
A:Cross-references: EMBL:AL031018
A:Experimental source: cultivar Columbia; BAC clone F7H19
C:Genetics:
A:Map position: 4
A:Introns: 140/3; 181/3; 216/3; 297/3; 362/3; 396/3
A:Note: F7H19.90
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	72.0%	Score 1839;	DB 2;	Length 444;
Best Local Similarity	78.7%	Pred. No. 3.9e-135;		
Matches 350;	Conservative 39;	Mismatches 40;	Indels 16;	Gaps 7;
QY	22	PEPSESLRHYSRMINSNHYTSP-----SRTIYSDRFITPSRSASKALFDIN-TPT-EGR 74		
DB	4	PVLSLES--RINRLIMNQSPSPSSLSRSIYSDRFITPSRSGNFALDIPSPSKDGK 61		
QY	75	DDSSAVYTLTLRTALFGDVAGPYTPKTDSPSMILPRNIRFKKTEKROSMHLSPFMD 134		
DB	62	EDDGSVATLTLRAAMFGE-----TPKRDITGRS--SRNIFRKTEKTHSLNSFSPGV 115		
QY	135	DDFVGVNHS-PVKAAPKRVSPRYKVLDAFALODDFYLVNLDVMSHNVLAAGNCVYLW 193		
DB	116	DDDSGVHSGPVAKPRKVPSPKVLDAFALODDFYLVNLDVMSHNVLAAGNCVYLW 175		
QY	194	NACSSKVRKLDLGVDCVCSVGMARCTHLAVGTNNGKVOIWAACRCKTRSMEGHRLR 253		
DB	176	NACSSKVRKLDLGAEDSVCSVGMALRGTHLAVGTSTGKVOIWDASRCKTRRTEGHRLR 235		
QY	254	VGALAWSSILSSGGRDNITVORIDRFVSKSGKRSVCGKMSYDRELASGND 313		
DB	236	VGALAWSSVLLSSGGRDSTLQDRDRCQEDVSKLAGKSEYCGKMSYDRELASGND 295		
QY	314	NKLFVWNGHSTQPVLYKCEHTAAVKAIAWSPHQLLASGGGTADRCIRFNNTTNSHLS 373		
DB	296	NKLFVWNGHSTQPVLYKCEHTAAVKAIAWSPHQLLASGGGTADRCIRFNNTTNSHLS 355		
QY	374	CMDTGSQVCNLMWSKNVDELSTHGYSONQIITWRYPIFMSKATLTGTTYVVLALASPD 433		
DB	356	SDTCSQVCNLMWSKNVDELSTHGYSONQIITWRYPIFMSKATLTGTTYVVLALASPD 415		

A:Cross-references: EMBL:AB005589; PIDN:BAA22152.1
 C:Genetics:
 A:Gene: str1: SPAC14.13c
 A:Map position: 1
 C:Function:
 A:Description: may control differentiation and cell cycling by negatively regulating Ctg
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 C:Keywords: cell cycle control

Query Match 41.7%; Score 1066; DB 2; Length 556;
 Best Local Similarity 45.1%; Pred. No. 6.4e-75;
 Matches 223; Conservative 73; Mismatches 121; Indels 78; Gaps 10;

QY 42 TSPS-----RTYSDREFIPSRASAKFALDINTPTG-----RDDSSAYTTLRLTAL 89
 DB 61 SSPNNKRRRTREGDRFIPSRDAST-ELMTGFTKVEGPLEVKKKQSVADRNFRTLLRSEL 119
 QY 90 FG-----PDVAGPVTEKTD-----PSMT-----LPN 112
 DB 120 FGSNDEFNNSPIATPNTTIGVSPRTDSDIDIELTQRTPPSSSHSSILQNTPTPS 179
 QY 113 RUIF-----RYKETROSMSLSLSPMDDEVPVGNHSPVKAP---RK 151
 DB 180 KRIFFYFLSPROKNSYCKKAOYQDNPRRTYSLP-----VRSITKDLISARLEGRE 233
 QY 152 VPRSPYKVLADALODDFYLLVDMSSHNVLAVGNCVYLMNACSSKVKIKDGLGVDC 211
 DB 234 LPSIFPRYLDAGLADDFYLLVDMSSHNVLAVGNCVYLMNACSSKVKIKDGLGVDC 211
 QY 212 VCSVGAOAGTGLAVGTNNKVOIWDARCKKIRSMESGRLVAGALAMSSLLSSGGRD 271
 DB 294 VSLKRVQGTGLAVGTNNKVOIWDARCKKIRSMESGRLVAGALAMSSLLSSGGRD 271
 QY 272 NIYQDRIQTQEDFVSKLSGKSEVCGLKMSYDNRELAAGNDKLFVWNOHSTQPVLYKC 331
 DB 354 HILHRDVRAPREHYFVFLTAHROEVCGLFEMNSNENLLASGGNDKLFVWNOHSTQPVLYKC 331
 QY 332 EHFAAVKATAMSPHLHGLASGGTADRCIRFNNNTTNSHLSCMDTGSQVONLVMSKNV 391
 DB 414 NHTAAVKAITWSPHOGILASGGTADRCIRFNNNTTNSHLSCMDTGSQVONLVMSKNV 391
 QY 392 ELVSTHGYSONOITVWRPYTMSKLAITLGHYRVLYLAISPGQITVYAGDETLRFNNV 451
 DB 474 EFLSTHGYSONOITVWRPYTMSKLAITLGHYRVLYLAISPGQITVYAGDETLRFNNV 451
 QY 452 FPPSKSNTSELSGA 466
 DB 534 FDS-KSKHSASTWS 547

RESULT 5

T40614
 G:Beta repeat protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
 C:Accession: T40614
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, December 1998
 A:Reference number: Z21941
 A:Accession: T40614
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-421 <LYN>
 A:Cross-references: EMBL:AL034563; PIDN:CAA22522.1; GSPDB:GN00067; SPDB:SPBC660.02
 A:Experimental source: strain 972h-; cosmid c660
 C:Genetics:
 A:Gene: SPDB:SPBC660.02
 A:Map position: 2
 A:Introns: 41/1
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 38.6%; Score 985; DB 2; Length 421;
 Best Local Similarity 45.5%; Pred. No. 8.5e-69;
 Matches 196; Conservative 68; Mismatches 131; Indels 36; Gaps 7;

QY 51 DRFIPSRAS-----KRALF-----DINPTGRODSSAYTTLRLTALFGPDVAG 96
 DB 3 DRFIPSRAS-----KRALF-----DINPTGRODSSAYTTLRLTALFGPDVAG 96
 QY 97 P-----VTPEKTDSPMTLPNNIFRYKTETROSMSLSLSPMDDEVPVGNHSPVKAP 150
 DB 61 SRATYGEDRKRIEKKMLDIPDRKSY-----SLSPISQSD-----MLRQPKPR 107
 QY 151 KVPSPYKVLADALODDFYLLVDMSSHNVLAVGNCVYLMNACSSKVKIKDGLGVDC 210
 DB 108 AFKTPYKILDAFYLKNDFFYLLVDMSSHNVLAVGNCVYLMNACSSKVKIKDGLGVDC 210
 QY 211 CVCSVGAOAGTGLAVGTNNKVOIWDARCKKIRSMESGRLVAGALAMSSLLSSGGRD 270
 DB 168 HVTSLMTGKGTGLAVGTNNKVOIWDARCKKIRSMESGRLVAGALAMSSLLSSGGRD 270
 QY 271 NIYQDRIQTQEDFVSKLSGKSEVCGLKMSYDNRELAAGNDKLFVWNOHSTQPVLYKC 330
 DB 228 EYTLHHDLRAPGCCAEMKKEHVEOEICGLQMDRSJGLASGGNDKLFVWNOHSTQPVLYKC 330
 QY 331 CERTAAVKAITAMSPHLHGLASGGTADRCIRFNNNTTNSHLSCMDTGSQVONLVMSKNV 390
 DB 288 EHFAAVKATAMSPHLHGLASGGTADRCIRFNNNTTNSHLSCMDTGSQVONLVMSKNV 390
 QY 391 NELVSTHGYSONOITVWRPYTMSKLAITLGHYRVLYLAISPGQITVYAGDETLRFNNV 450
 DB 348 NELVSTHGYSONOITVWRPYTMSKLAITLGHYRVLYLAISPGQITVYAGDETLRFNNV 450
 QY 451 VF-PSPKSN 460
 DB 408 LENKRPKEEST 418

RESULT 6

S64005
 hypothetical protein YGL003c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G3765
 C:Species: Saccharomyces cerevisiae
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 26-May-2000
 C:Accession: S64005
 R:Hehling, U.; Hofmann, B.; Delius, H.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64003
 A:Accession: S64005
 A:Molecule type: DNA
 A:Residues: 1-566 <HEB>
 A:Cross-references: EMBL:Z72525; NID:q1322450; PIDN:CAA96703.1; PID:e243314; PID:q132
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:CDH1
 A:Cross-references: SGD:S0002971; MIPS:YGL003c
 A:Map position: 7L
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 36.5%; Score 933.5; DB 2; Length 566;
 Best Local Similarity 37.5%; Pred. No. 1.3e-64;
 Matches 209; Conservative 82; Mismatches 153; Indels 113; Gaps 13;

QY 6 NRPNPSTVADNSPPPEP--SPESLKHYSRMINSNH---YTSPEP-----TIYSDRPI 54
 DB 4 NLNP-----FMNNTPPSSPLKGSSEKRSKRPISSSSSASLSSPSKRSPSTVYGDRIY 58
 QY 55 PSR-----SASKALPDINTPTGGRD-----DSSAYTTLRLTALFG----- 91
 DB 59 PSRTDIDFNSIVSISMAVPAALNPSTFEDQVYKQKRAHETRYNTLLKNELFGEMLSKD 118
 QY 92 -----PDVAG-----PYTPE 101

```

Db 119 TVGSSSIDRLKINRPSRGVNAHENTTRHGYELERSTPPPEAGLEEFSPHSTVPR 178
QY 102 K---TDSPSMTLPNNIR-----YKTE-----TROSMHSLSPMDDEFPVGV 141
Db 179 RLFTSQDEITRPSNSNRGASLLTYQQRKRRLSASLOSOFEDSPSPVPPS--KQL 236
QY 142 NHPYKARPRSPRYKVLADAPALQDDFYLNLVDMSHNVAVLAVGNCVYLMNACSSKYT 201
Db 237 LLSPEKQFQIAKPYRYVLADPRLADDEYYSLIDMSSSDVLAVALGSLFILDNTDGV 296
QY 202 KLCDDGVDDCYGVMQAGTHLAVGNNKVOIWDARCKKIRSMGHLRGALAMSS 261
Db 297 HLCD--TENETYSLSMIGAGSHLAVGQANLVEIYDWMKRCIPTLSGHIDRVACLSMN 354
QY 262 SLSSGGDKNIYQDIRTQDEYFSKLSGHSKSEYGLKWSYDNEELASGGNDKLPVWQ 321
Db 355 HVLTSGRDHRILHRDVRMPDPFETIESHQEVCGIKMVAADKGLASGGNDVNVHYEG 414
QY 322 HSTQPVLYKCEHTAAVAIAVMSPHLGLASGGGTADRCIRFNTTNSHLSGMDTGSQV 381
Db 415 TSKSPILTFDEHKAAYAMAMSPHKGVLATGGTADRLKINNVMSIKMSIDSGSI 474
QY 382 CNLWMSKNVELYSTHGYSONOIIWRYPMKSLATLGHTRYVLYLAISPDGQITVGA 441
Db 475 CNMWSKNTMELVTSRYSKYNTLMDCNMDPLAILKHSFVHLTLSDGTIVVSGA 534
QY 442 GDETLRFWNVFSPKSO 458
Db 535 GDETLRYWKLDFDKPAK 551

```

RESULT 7

```

T05977
hypothetical protein F17M5.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-May-2000
C:Accession: T05977
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hobeisel, J.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15263
A:Accession: T05977
A:Molecule type: DNA
A:Residues: 1-447 <BEV>
A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.20
A:Experimental source: cultivar Columbia; BAC clone F17M5
C:Genetics:
A:Gene: ATSP:F17M5.20
A:Map position: 4
A:Introns: 23/3; 116/3; 206/3; 400/3
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

```

```

Query Match 36.1%; Score 923; DB 2; Length 447;
Best Local Similarity 41.1%; Pred. No. 6; Le-64;
Matches 196; Conservative 87; Mismatches 124; Indels 70; Gaps 11;

```

```

QY 1 MDGGRNRPPTSTVRDMSPPESPESLRAVSRMINSNHTSPRTIYSDREFIPSSAS 60
Db 1 MDAGLRCP-----LQEHFLPRKNSKVL-----DREFIPNSAM 34
QY 61 KF--ALPDIPTBGRDSSAVYTLTLTALEFPDVAGPVP--EKTDSPSMTLPNNIR 116
Db 35 NEDVAHFALTEERKQDQASAT-----VSSPSKEAYRKQLAETMNLNHTRL 80
QY 117 RYKTETRSMHSLSPMDDEFPVGNHS-----PYKAPRKVPSPRYKVLADAPALOD 168
Db 81 AFRKK-----POAPVELLP--NHSASLHQOPKSKVPRRYIPQTSERTLAPDITVD 130
QY 169 FYLLVDWSSHNVLAVGLGNCVYLMNACSSKYTKLCDLGYD--DCVCSVGAQGTILAVG 227
Db 131 FYLLVDWSSHNVLAVGLGNCVYLMNACSSKYTKLCDLGYD--DCVCSVGAQGTILAVG 227
QY 228 TNNCKVOIWDARCKKIRSMGHLRGALAMSSSLSSGGDKNIYQDIRTQDEYFS 286

```

```

Db 191 LNSSEYQIMDSASNRQLTKLKGHSRVSGLAMNNHILTTGGMDGLINNDRIIRSPIYE 250
QY 287 KLSGHSKSEYGLKWSYDNEELASGGNDKLPVWQ-----HSTQPVLYKCEHTAAVAI 340
Db 251 TYRGTQEVCGILKWSGSGQOLASGGNDVNVHMDRSVASSNSTQWLHLRLEHTSAVKAL 310
QY 341 AMSPHLGLASGGGTADRCIRFNTTNSHLSGMDTGSQVGNLWMSKNVELYSTHGS 400
Db 311 AMCPQANILATGGGGRDRTIKFMNHTGACLSNVDTSQVCSLMSKNERELSSHGFT 370
QY 401 QNQLIYWRPYSKSLATLGHTRYVLYLAISPDGQITVGADETLRFWNVFSPKS 457
Db 371 QNQLIYWRPYSKSLATLGHTRYVLYLAISPDGQITVGADETLRFWNVFSPKS 457

```

RESULT 8

```

T05978
hypothetical protein F17M5.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 26-May-2000
C:Accession: T05978
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hobeisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15263
A:Accession: T05978
A:Molecule type: DNA
A:Residues: 1-457 <BEV>
A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.30
A:Experimental source: cultivar Columbia; BAC clone F17M5
C:Genetics:
A:Gene: ATSP:F17M5.30
A:Map position: 4
A:Introns: 33/3; 126/3; 216/3; 410/3
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

```

```

Query Match 36.0%; Score 920.5; DB 2; Length 457;
Best Local Similarity 41.8%; Pred. No. 9; Le-64;
Matches 191; Conservative 84; Mismatches 117; Indels 65; Gaps 10;

```

```

QY 21 PPEPSPESLRAVSRMINSNHTSPRTIYSDREFIPSSASKF--ALPDINTPREGDRSS 78
Db 26 PRKPSKENVL-----DREFIPNSAMNPFYAHFALTEGRKQDQTA 64
QY 79 SAYTTLRTALFEPDVAGPVP--EKTDSPSMTLPNNIRIYRTETROSMSLSPPMDDO 136
Db 65 A-----VSSPSKEAYRKQLAETMNLNHTRLAIFRNK-----POAPVE 101
QY 137 FVPGVNS-----PYKAPRKVPSPRYKVLADAPALQDDFYLNLVDWSSHNVLAVGLGN 188
Db 102 LKPS--NHSASLHQOPKSKVPRRYIPQTSERTLAPDITVDDEFYLLNDWGSANVLAIALDH 160
QY 189 CYLLMNAKSSKYTKLCDLGYD--DCVCSVGAQGTILAVGNNKVOIWDARCKKIRSM 247
Db 161 TVYLMADSTSTSELVITIDEKGPVTSINMAPGRHVAAGLNSEYQIMDSASNRQLRTL 220
QY 248 E--GHLRVGALAMSSSLSSGGDKNIYQDIRTQDEYFSKLSGHSKSEYGLKWSYDNE 306
Db 221 KGGHQRVSGLAMNNHILTTGGMDGLINNDVIRISPIYETRGHTQEVCGILKWSGSGQO 280
QY 307 LASGGNDKLPVWQ-----HSTQPVLYKCEHTAAVAIAVMSPHLGLASGGGTADRC 360
Db 281 LASGGNDVNVHMDRSVASSNSTQWLHLRLEHTSAVKALAMCPQANILATGGGGRRT 340
QY 361 IRFWMNTTNSHLSGMDTGSQVGNLWMSKNVELYSTHGSQOIIWRYPMKSLATLITG 420
Db 341 IKFMNHTGACLSNVDTSQVCSLMSKNERELSSHGTOQLTMKYPYKVAELTG 400
QY 421 HTRYVLYLAISPDGQITVGADETLRFWNVFSPKS 457
Db 401 HTRYVLYLAISPDGQITVGADETLRFWNVFSPKS 457

```

Query Match	35.3%;	Score 901.5;	DB 2;	Length 420;
Best Local Similarity	43.1%;	Pred. No. 2.9e-62;		
Matches 196; Conservative	72;	Mismatches 124;	Indels 63;	Gaps 11

RESULT 10

C;Superfamily: unassigned WD repeat proteins; WD repeat homology
F;305-338/Domain: WD repeat homology <WD1>

RESULT 11

Query M

Best Local Simultaneously 81; Mismatches 165; Indels 53; gaps
Matches 187; Conservative -----MINSNHTSPERTIYSDRFI 54
QY 8 NPPPTSTVRDNDSPPEPSPESLHVS-----

Db 22 NAPPARWOKKAKKAAGPAPSPMRAANSHSAGRTPEKTPKSSSKVOTTSKP-GGDRY 80
 QY 55 PPSRSKFPALPDINTPTTEGRDSSAYTTTLRTALFGPDVAGPYTEKTPSPMTLPNN 114
 Db 81 PHSRAQMEVASFLLSKENQSENQ-----PTKKHOKAALNIN 121
 QY 115 -----IFRYKTETROS-----MHSLSPEMDDEFGVNHSPVKAPKPYKAVDA 162
 Db 122 GFDEVEAKIILRLSGKPPONAEQYONRLKVLYSOKATFG-----SRKTCRYTSLPRLIDA 178
 QY 163 PALODDEYLYNDVMSHNNVAVLGNCLYIMANASCKVTKLCDL-GVDDCVCSVGAORG 221
 Db 179 PEIRNDVYLYLVMSHNNVAVLGNCLYIMANASCKVTKLCDL-GVDDCVCSVGAORG 221
 QY 222 THAVGTNNKQVQIWDAAKCKIRSMEGHRLRYGALAMSSLLSGGDKNIYODITQ 281
 Db 239 NYLAVGTSSAEVQVQIWDAAKCKIRSMEGHRLRYGALAMSSLLSGGDKNIYODITQ 281
 QY 282 EDEFSKLSGHSVEGGLKMSYDNRELASGDNKLFV-----NOSTOPVLYKCEHTAAV 337
 Db 299 EHHVATLSGHSVEGGLKMSYDNRELASGDNKLFV-----NOSTOPVLYKCEHTAAV 337
 QY 338 KAIAMSPHLJGLASGGTADRCIRFMTTNSHLSGMDTSGOYCNLYMSKNNVELYTH 397
 Db 359 KAVAMCPMOSNVLATGGTSDRHRIMVWCGACLSAVDAHSQVCSILMSPHYKELISGH 418
 QY 398 GYSQNOIYVRYPTMSKLTATLGHTRYLYLAISPDGQITVGADETLRFVWVFPSPKS 457
 Db 419 GYNOQVLYVRYPTMSKLTATLGHTRYLYLAISPDGQITVGADETLRFVWVFPSPKS 457
 QY 458 QNTSE 463
 Db 479 RRRERE 484

RESULT 12

T01768

hypothetical protein A_IG002P16.8 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 26-May-2000
 C:Accession: T01768

R:Miller, N.; Beck, C.; Kramer, J.
 Submitted to the EMBL Data Library, June 1997
 A:Description: The sequence of A. thaliana IG002P16.
 A:Reference number: Z14421

A:Accession: T01768
 A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-440 <MIL>

A:Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191163; GSPDB:GN00063; ATSP:A_IG002P16.8

A:Gene: ATSP:A_IG002P16.8
 A:Map position: 5

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 34.3%; Score 876; DB 2; Length 440;
 Best Local Similarity 40.6%; Pred. No. 2.7e-60;
 Matches 189; Conservative 74; Mismatches 129; Indels 74; Gaps 10;

QY 12 TSTVDNSPPPEPESLRLHVSRLMNSHNTSPRTIYSDRTSPRSASFALPDINTPT 71
 Db 4 TCTVPEHFLPKRLSKONT-----DRTIPNRSAKQFD-FANYALT 41
 QY 72 EG-----RDDSSAYTTTLRTALFGPDVAGPYTEKTPSPMTLPNN-----IFRY 118
 Db 42 QGSKRNDEVTASAKKAYTQALAVM-----NONRRIIAFNN 79
 QY 119 KTETRQSK-HSLSPFMDDEFGVNHSPVKAPKPYKAVDAALODDEYLYLVMS 177
 Db 80 KPKSLSTNHSDSPHQ-----NRPVKAPRYTIPONSERVLDAAGLDRDESLMLDMG 131

QY 178 SHNVLAVGLNCVYLYLMANASCKVTKLCDLVD-DCVCSVGAORGTHAVGTNNKQVQI 236
 Db 132 SANVLALALGDTYLYLMANASCKVTKLCDLVD-DCVCSVGAORGTHAVGTNNKQVQI 191
 QY 237 DAARCKIRSMW-CHRLRYGALAMSSLLSGGDKNIYODITQIODEVSKLSGHKSEV 295
 Db 192 DCVSNQVRLRLRGHESRVSGLAMDNHLLTGTGMDKRYVNNQVRLRSIVETYLGHTEEV 251
 QY 296 CGLKWSDNRELASGDNKLFV-----NOSTOPVLYKCEHTAAVAKIAMSPLHGL 349
 Db 252 CGLKWSDNRELASGDNKLFV-----NOSTOPVLYKCEHTAAVAKIAMSPLHGL 349
 QY 350 LASGGTADRCIRFMTTNSHLSGMDTSGOYCNLYMSKNNVELYTHGYSQNOIYVRY 409
 Db 312 LATGGVGVQOKIKFMTTNSHLSGMDTSGOYCNLYMSKNNVELYTHGYSQNOIYVRY 371
 QY 410 PTMSKLTATLGHTRYLYLAISPDGQITVGADETLRFVWVFPSP 455
 Db 372 PMSKMAELNGHTSHRYLFAOQSPNCTVASAAGDENLRLMVFGEF 417

RESULT 13

T41719

CDG20 protein homolog - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
 C:Accession: T41719

R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 Submitted to the EMBL Data Library, October 1999
 A:Reference number: Z22012

A:Accession: T41719
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-488 <RIE>

A:Cross-references: EMBL:AL121770; PID:CA557442.1; GSPDB:GN00066; SPDB:SPAC821.08c

A:Experimental source: strain 972h; cosmid c821

A:Gene: SPDB:SPAC821.08c
 A:Map position: 1

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 30.4%; Score 777.5; DB 2; Length 488;
 Best Local Similarity 37.4%; Pred. No. 1.4e-52;
 Matches 181; Conservative 73; Mismatches 169; Indels 61; Gaps 12;

QY 2 DGTGNPNPPTSTVRDSSPPPE-----PESLRLHVSRLMNSHNTSPRTIYSDRF 53
 Db 42 NGRSSKRCSPKSSFTIRNSPKIDVNTDWSIPLCGSPRKRKSR-----PASR---SDRF 90
 QY 54 IPSR--SASKFALPDINTPTTEGRDSSAYTTTLRTALFGPDVAGPYTEKTPSPMTLP 111
 Db 91 IPSRPTANAF-----VNSIS---SDVPPDISEVAEAC-GEFLNTRVLAKFLAPKPKP 142
 QY 112 NNIRFRYKTEQOSMHSLSPEMDDEFGVNHSPVKAP--RKYVRSPPYKVLDAALODDE 169
 Db 143 ---VDLRTQHNRRPQ-----PVYPAKRRENTPTPERVLDAAGLIDDI 181
 QY 170 YLNLVDMSSHNVAVLGNCLYLYLMANASCKVTKLCDLGVDDCVCSVGAORGTHAVGTN 229
 Db 182 YLNLDMSNLNVAVLGNCLYLYLMANASCKVTKLCDLGVDDCVCSVGAORGTHAVGTN 229
 QY 230 NGKVOIWDAAKCKIRSMEGHRLRYGALAMSSLLSGGDKNIYODITQIODEVSKLS 289
 Db 242 NGVDIVDYESQVTKLRTMAGHQAQVGLSNRRHVLVSSGSRSAIHHHVDRIANHOIGTLQ 301
 QY 290 GHRSEVCGLKWSDNRELASGDNKLFVNNHSTOPLVLYKCEHTAAVAKIAMSPLHGL 349
 Db 302 CHSEVCGLKWSDNRELASGDNKLFVNNHSTOPLVLYKCEHTAAVAKIAMSPLHGL 349
 QY 350 LASGGTADRCIRFMTTNSHLSGMDTSGOYCNLYMSKNNVELYTHGYSQNOIYVRY 409
 Db 362 LATGGTMDKQIHFMNAATGARVNTVDASQVTLMSPHSKREIMSTHGFPDNNLSIWY 421

Db 276 AIALDT

Wed May 22 11:09:48 2002

Search completed: May 21, 2002, 11:46:48
Job time: 156 sec

us-09-701-572-2.rpt

On process:		May 21, 2002, 11:46:52 ; Search time 13.47 Seconds
Run on:	(without alignments)	
	sec 300 million cells	updates/sec

```

1000      2555
Perfect score: 1 MDGTCNRNPPTSTVNDSP.....KSONTESEIGALSGRTTIR 4/3
Sequence:

```

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 100222

Minimum	DB seq	length:	0
Maximum	DB seq	length:	200000000000

Post-processing:	Minimum Match 0%
	Maximum Match 100%
Listing first	45 summaries

Database : SwissProt_40:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB	ID	Description
1	933.5	36.5	566	1	YGA3 YEAST
2	674	26.4	610	1	CC20 YEAST
3	431	16.9	535	1	YA98_SCHPO
4	290	11.4	742	1	PKM4_THECU
5	273	10.7	1258	1	YS00_AANSP
6	272	10.6	1356	1	HEP1_PODAN
7	254.5	10.0	614	1	TU11_SCHPO
8	250	9.8	586	1	TU12_SCHPO
9	237	9.3	515	1	YCW2_YEAST
10	221	8.6	514	1	TUPL_CANAAL
11	217	8.5	604	1	RCOI_NECR
12	209	8.2	486	1	PELI_ARATH
13	206.5	8.1	339	1	CIAL_HUMAN
14	203	7.9	465	1	PR04_YEAST
15	200	7.8	409	1	YG45_YEAST
16	197	7.7	682	1	TUPL_KILUA
17	196.5	7.7	361	1	WDS_DROME
18	195	7.6	340	1	GBI1_HOMAN
19	194	7.6	606	1	PF20_CHLME
20	193	7.6	318	1	PEX7_MOUSE
21	193	7.6	348	1	YE91_SYNS3
22	193	7.6	732	1	KMRB_DICDI
23	193	7.6	1047	1	HIRA_DROME
24	192.5	7.5	340	1	GBR2_RAT
25	192.5	7.5	925	1	PMW2_CAEEL
26	192	7.5	340	1	YGB1_CAEEL
27	192	7.5	309	1	GABE_SCHPO
28	190.5	7.5	608	1	WDRI_XENLA
29	189.5	7.4	314	1	GBR2_SCHPO
30	188.5	7.4	501	1	YH92_CAEEL
31	188.5	7.4	713	1	TUPL_YEAST
32	188	7.4	577	1	57L1_HUMAN
33	187.5	7.3	554	1	PR04_ARATH

45	185	7.2	409	1	WDRE_HUMAN
44	185	7.2	359	1	GBB_CRYPA
43	185	7.2	347	1	GBB_DICI
42	185	7.2	340	1	GBB_MOUSE
41	186	7.3	962	1	YA4E_SCHPO
40	186	7.3	608	1	WDRI_DROME
39	186	7.3	409	1	LISI_MOUSE
38	186	7.3	409	1	LISI_HUMAN
37	187	7.3	341	1	GBRI_DROME
36	187	7.3	341	1	GBRI_LYMSI
35	187	7.3	334	1	WDRE_MOUSE
34	187	7.3	334	1	WDRE_HUMAN
33	187	7.3	334	1	WDRE_MOUSE
32	187	7.3	334	1	WDRE_MOUSE
31	187	7.3	334	1	WDRE_MOUSE
30	187	7.3	334	1	WDRE_MOUSE
29	187	7.3	334	1	WDRE_MOUSE
28	187	7.3	334	1	WDRE_MOUSE
27	187	7.3	334	1	WDRE_MOUSE
26	187	7.3	334	1	WDRE_MOUSE
25	187	7.3	334	1	WDRE_MOUSE
24	187	7.3	334	1	WDRE_MOUSE
23	187	7.3	334	1	WDRE_MOUSE
22	187	7.3	334	1	WDRE_MOUSE
21	187	7.3	334	1	WDRE_MOUSE
20	187	7.3	334	1	WDRE_MOUSE
19	187	7.3	334	1	WDRE_MOUSE
18	187	7.3	334	1	WDRE_MOUSE
17	187	7.3	334	1	WDRE_MOUSE
16	187	7.3	334	1	WDRE_MOUSE
15	187	7.3	334	1	WDRE_MOUSE
14	187	7.3	334	1	WDRE_MOUSE
13	187	7.3	334	1	WDRE_MOUSE
12	187	7.3	334	1	WDRE_MOUSE
11	187	7.3	334	1	WDRE_MOUSE
10	187	7.3	334	1	WDRE_MOUSE
9	187	7.3	334	1	WDRE_MOUSE
8	187	7.3	334	1	WDRE_MOUSE
7	187	7.3	334	1	WDRE_MOUSE
6	187	7.3	334	1	WDRE_MOUSE
5	187	7.3	334	1	WDRE_MOUSE
4	187	7.3	334	1	WDRE_MOUSE
3	187	7.3	334	1	WDRE_MOUSE
2	187	7.3	334	1	WDRE_MOUSE
1	187	7.3	334	1	WDRE_MOUSE

ALIGNMENTS

```

SUITE
1
A3 YEAST STANDARD: PRT: 566 AA.
YGA3 YEAST
P5197;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical 62.8 kDa Trp-Asp repeats containing protein in PMc1-TFG2
intergenic region.
YGI003C.
Saccharomyces cerevisiae (Baker's yeast);
Saccharomycetes;
Fukariyola; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
(1)
SEQUENCE FROM N.A.
Hehling U., Hofmann B., Dalius H.;
Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
-1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAIN).
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z7525; CAA96703.1; -
SGB; S0002971; YGI003C.
DR InterPro: IPR000002; F1ZZY.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD004563; F1ZZY; 1.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS00682; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; POLY-SBR.
FT DOMAIN 32 38 WD 1.
FT REPEAT 300 330 WD 2.
FT REPEAT 342 370 WD 3.
FT REPEAT 383 413 WD 4.
FT REPEAT 425 458 WD 5.
FT REPEAT 469 501 WD 5.
FT REPEAT 513 543 WD 6.
FT REPEAT 566 AA; 62821 MW; DA6D5545607906B CRC64;
SEQUENCE

```

36.58; Score 933.5; DB 1; Length 566;

Query Match	37.5%;	Pred. No.	5.8e-66;	
Best Local Similarity				Indels 13;
Matches 209; Conservative	82;	Mismatches	153;	Gaps 13

QY 6 NRNRPPTSTVRDNSPPPEP--SPESGLHVRKMSINSH-----YTSPSR-----TLLSDNLT
||| :||| ||| | :||| :||| | :|||

CC	MICROTUBULES.
CC	- - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC	- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 502
CC	ONWARD AND IS SHORTER (519 AA) DUE TO A FRAMESHIFT.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaborator
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as their content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC	or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; X59428; CAA42058.1; -
DR	EMBL; Z72638; CAA96824.1; -
DR	EMBL; D16506; BAA03957.1; -
DR	PIR; S18840; S18840.
DR	SGD; S0003084; CDC20.
DR	InterPro; IPR000002; Flyzy.
DR	InterPro; IPR001680; WD40.
DR	Pfam; PF00400; WD40; 5.
DR	ProDom; PD004563; Flyzy; 1.
DR	SMART; SM00320; WD40; 5.
DR	PROSITE; PS00678; WD_REPEATS_1; 4.
DR	PROSITE; PS0082; WD_REPEATS_2; 2.
KW	PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW	Cell division; Microtubules; Repeat; WD repeat.
FT	REPEAT 257 .. 296
FT	REPEAT 299 .. 338
FT	REPEAT 342 .. 379
FT	REPEAT 383 .. 422
FT	REPEAT 425 .. 467
FT	REPEAT 469 .. 519
FT	REPEAT 523 .. 562
FT	CONFLICT 318 .. 319
FT	CONFLICT 502 .. 519
FT	CONFLICT 520 .. 610
SO	SEQUENCE 610 AA; 67359 MW; CCECD149C1F5ACF CRC64;

OY 413 SKLATLT-GHTYRLVLAISPDCQITVAGDETLFENNPPSKQTESIGAL 467
 DB 515 FKVAEVAHARICQSOLSPDGTTLATVGGDENLKFALF-DPROTGSRREDIGM 569

RESULT 3
 Y98_SCHPO STANDARD: PRT: 535 AA.
 ID Y98_SCHPO
 AC 009786;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 60.5 kDa Trp-Asp repeats containing protein C1366.08 in
 chromosome 1.
 GN SPAC1366.08.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RA Odell C., Bowman S., Bartell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
 CC -1 SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Z54308; CAA91101.1;
 CC InterPro: IPR000002; F1zzy.
 CC InterPro: IPR001680; WD40.
 CC Pfam: PF00400; WD40; 5.
 CC ProDom: PD004563; F1zzy; 1.
 CC ProDom: PD00320; WD40; 3.
 CC SMART: SM00320; WD_REPEATS_1; FALSE_NEG.
 CC PROSITE: PS00678; WD_REPEATS_1; FALSE_NEG.
 CC PROSITE: PS00678; WD_REPEATS_2; 2.
 CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
 CC DR Hypothetical protein; Repeat: WD repeat.
 CC KW REPEAT
 CC FT REPEAT 189 226 WD 1.
 CC FT REPEAT 228 267 WD 2.
 CC FT REPEAT 269 314 WD 3.
 CC FT REPEAT 320 359 WD 4.
 CC FT REPEAT 362 404 WD 5.
 CC FT REPEAT 462 505 WD 6.
 CC FT REPEAT 535 AA; 60533 MW; F438AE5981C65B42 CRC64;
 SO SEQUENCE

Query Match 16.9%; Score 431; DB 1; Length 535;
 Best Local Similarity 27.8%; Pred. No. 2e-26; Indels 112; Gaps 20;
 Matches 142; Conservative 80; Mismatches 176;

OY 12 TSTV-----RDNSPP-PEPSPESLRHYSRMINSNHYTSPRTIYS-----DRF 53
 DB 30 SSTVYGMNERSTPLTIDPSNKEFANYC-PIKSKLFGSYKSGYRSQRLTSRNGIDR 88
 OY 54 IPSASAKFALPDINTPEGRDS-----SAYTTLRLALFGPDVAGPVTPPE 101
 DB 89 IPMTS-----NKDTISGRHSLSRLNKNKASSETYOOLLEYAL-----EVE 132
 OY 102 KTDSPSKTLPRNITFKKTETROSMSLSPPM---DDDFVPGVNSPVKAPKYPKSPKX 158
 DB 133 RDD-----NVTYAKILOKSDMKOCTPVASEKDKGKLNKKNKSPENL--LPER 181
 OY 159 VLDAPALODDFYLNLDVMSHNVLAVGLGNCVYLNNACSKYTKLCCDDGVDDCVCSYQMA 218
 DB 182 IIDAELRDRDFTYSLTSGWSPKGLAIGLAENITLWSK-ELGPRVLEESITV-VSSVAYS 239

OY 219 ORGTHLANGTNGKVOIN-DARCKKIRSMGHRRLRYGALAMSSSL-----LSGGGRKNI 273
 DB 240 YNGDILAVGRVDTGLQFQODNERPRII--HHPGDIGVLAHMYETLNLLGKGNGLNI 297
 OY 274 YORDIRTOEDFYSK-----LSGKSEYCGLKSYNDRELASGDNKLFVNNHSTOP 326
 DB 298 FYVDIITSES-TSKAVLVATITNAHDEQVCGLTWNHDSOPASGDNRRVCLFSGDLRG 356
 OY 327 VLKXCEHTAAVKAIAHSPHGLHGLASGGTADRCIRFNTTNSHLSCMDTGSQVCLVW 386
 DB 357 PLYWQDQNAVVKLSFCPPQORSILATGAGSHKHIRFYNCNGKKIDELVCGAQITSIHW 416
 OY 387 SKNNNELVSTHGSY---ONQIIVWRPYPMKSLATLTGHTYVLYLAISP- 433
 DB 417 SPKRYEVSYVFGYSLETVQHRFAVYSMPOLCVSV-----LPEVPDIRCVHSVLT 467
 OY 434 -----GGTITVAGDETLFENN 451
 DB 468 SOLNETTGRCMSTDSIITASNETIKFEDL 497

RESULT 4
 PKWA_THECU STANDARD: PRT: 742 AA.
 ID PKWA_THECU
 AC P49695;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative serine/threonine-protein kinase PKWA (EC 2.7.1.-).
 GN PKWA OR PKW1.
 OS Thermomonospora curvata.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporanginae; Thermomonosporaceae;
 OC Thermomonospora.
 CC NCBI_TaxID=2020;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CCM 3352;
 RA Janda L., Tichy P., Spizke J., Petrlick M.;
 RL MEDLINE=96200125; PubMed=86317732;
 RT "A deduced Thermomonospora curvata protein containing
 RT serine/threonine protein kinase and WD-repeat domains";
 RT J. Bacteriol. 178:1487-1489(1996).
 CC -1 FUNCTION: MAY PLAY A REGULATORY ROLE DURING THE COMPLEX GROWTH
 CC CYCLE AND IN SECONDARY METABOLITE PRODUCTION.
 CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1 SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF115133; AAB05822.1;
 CC HSPSP, P00523; 2PTR.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002290; Ser_thr_pkinase.
 CC InterPro: IPR001680; WD40.
 CC Pfam: PF00400; pkinase; 1.
 CC Pfam: PF00400; WD40; 7.
 CC PRINTS: P00320; GPOTEINBRPT.
 CC SMART: SM00320; WD40; 7.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE: PS00678; WD_REPEATS_1; 5.
 CC PROSITE: PS50082; WD_REPEATS_2; 7.
 CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
 CC Transferase: Serine/threonine-protein kinase; ATP-binding; Repeat;

KM WD repeat. 16 266 PROTEIN KINASE.
 FT DOMAIN 22 30 ATP (BY SIMILARITY).
 FT NP_BIND 22 30 ATP (BY SIMILARITY).
 FT BINDING 44 44
 FT ACT_SITE 138 138
 FT DOMAIN 301 394
 FT REPEAT 455 496
 FT REPEAT 497 538
 FT REPEAT 539 580
 FT REPEAT 581 621
 FT REPEAT 622 663
 FT REPEAT 664 705
 FT REPEAT 706 742
 SO SEQUENCE 742 AA: 78950 MW: AC173460DB4383D CRC64;

Query Match 11.4%; Score 290; DB 1; Length 742;
 Best Local Similarity 25.7%; Pred. No. 3.9e-13;
 Matches 123; Conservative 71; Mismatches 199; Indels 86; Gaps 20;

QY 12 TSTVRDNPSPPEPSRLHVSRLMINS---NHYSRSRTYSDRFPSRSASKFALFDI 67
 DB 266 THHTGDWPP-----HLSDLGSMPLGATTSNPPLALEPPPSHGPPR----- 311
 QY 68 NPTREGDDSSAYTTLRTALFGPDVAGVPTPEKTDSPMTLPNNRIERYKTETROSMH 127
 DB 312 --PSEPLDP-----GDDADP-----SAKPSRTLPPEPEPELEKPIQVIH 352
 QY 128 S---LSPMDDDFVGVNHSVPKPKRVKLVLDALDDDFVNLVWSS-----H 179
 DB 353 EPERAPTPRRREPARGATKPKNPR--PAAPQPPWSPRPVQPPRMKQILTKKPAAGILT 410
 QY 180 NVLAVGNCVYLMN-----ACSSKVTLCID-----LGVDDCVCSVGMQR 220
 DB 411 AVATAGLVSVLWQWLTPEPTPLRPDSTAPESADPHNLNPRLLTDRRAVAVAFSPG 470
 QY 221 GTHLAVGTNNCKVOIMDAARCKIRSMGHRIRVGLAMS--SLLSSGGRDKNIYORDI 278
 DB 471 GSLAGSGDGLIHWDVASCDELHTLBSGHTDWRVAVAFSPDGLIAGSDDATVRLMDV 530
 QY 279 RTEDEPVSKLSGKSEVCCILKWSYNNRELASGNDNKFVWN-----OHSPVYLKCEH 333
 DB 531 AAEE-RAVEGHTHYVDIAFSPDGSVAVASGRGTARLNVATGEHA---VLK--GH 584
 QY 334 TAAVAIAIAPHLHGLASGGGTADRCIRFNNTTNSHSCMDTGSQ--VCNLVWSKNVE 392
 DB 585 TTYVAVAFSPD--GSMVAS--GSRDGTIRLMDVATGKERDYLQAPENNVSLAFSPDGS 641
 QY 393 LVSTGYSONOIIVWRYPTNSKLAITLGHYRVLYLAISPDQTIYTGAGDETLRFNMY 451
 DB 642 LV--HG-SDSYVHMDVAVAGEALHTFEHGDWVRAVAFSPDGLIAGSDSDTIRLMDV 697

RESULT 5
 YSOO_ANASP STANDARD; PRT; 1258 AA.
 AC Q8YTC2;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE Hypothetical WD-repeat protein alr2800.
 GN ALR2800.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Matsumoto A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;

"Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120.".
 Res. 8:203-213(2001).
 --- SIMILARITY: CONTAINS 15 WD REPEATS (TRP-ASP DOMAINS).
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AP003590; BAB74499.1; .
 DR PROSITE; PS00678; WD_REPEATS.1; 9.
 DR PROSITE; PS50082; WD_REPEATS.2; 14.
 DR PROSITE; PS50294; WD_REPEATS_REGION.1.
 KW Hypothetical protein: Repeat; WD repeat; Complete proteome.
 FT REPEAT 55 93
 FT REPEAT 640 679
 FT REPEAT 682 721
 FT REPEAT 724 763
 FT REPEAT 766 807
 FT REPEAT 809 849
 FT REPEAT 850 889
 FT REPEAT 892 931
 FT REPEAT 934 975
 FT REPEAT 976 1017
 FT REPEAT 1019 1059
 FT REPEAT 1060 1101
 FT REPEAT 1103 1143
 FT REPEAT 1144 1183
 FT REPEAT 1186 1227
 SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;

Query Match 10.7%; Score 273; DB 1; Length 1258;
 Best Local Similarity 29.9%; Pred. No. 1.8e-13;
 Matches 81; Conservative 54; Mismatches 126; Indels 10; Gaps 8;

QY 184 VGLNCYCYLLNAGSKVTKLCIDGVDCVSGVMQRGTHLAVGNCKKVOIMDAARCK 243
 DB 912 VSLDQSVRLNCRGTGCKLAK--YGNIDMALPYAFSPDQILASGNSKRYKVLMDMQRTK 970
 QY 244 IRSMEGHRIRLVGLAMS--SLLSSGGRDKNIYORDIFTOEDFVSKLSGKSEVCGLKS 301
 DB 971 ISLEGHTEPIYGIASFSPDQTLASASTDSSVRIMNISTGCFQILDE-HTDWYAVAFH 1029
 QY 302 YDNRELASGNDNKLFWNQHSTQVLEKCEHTAAVAKAIAHSPHLHGLASGGGTADRCI 361
 DB 1030 PQCKIATGSACTVAKLWNISITGCKLTKLSHSDKIIGMAMSPD--GOLLA--ASADQSV 1086
 QY 362 RFWNNTTNSHLSM--DTGSOVCNLWMSKNNEVSTHGYSQNOIIVWRYPTMSKLATLTG 420
 DB 1087 RLMDCTGCGVILRGHNRYSALFSPN--GEIATCTSDOT--VKIMDQGGCKLKTILG 1144
 QY 421 HTYRVLYLAISPDQTIYTGAGDETLRFNMY 451
 DB 1145 HTNWVFDIAFSPDGLIAGSDSDTIRLMDV 1175

RESULT 6
 HET1_PODAN STANDARD; PRT; 1356 AA.
 AC Q00808;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Vegetative incompatibility protein HET-1.
 GN HET-1.
 OS Podospora anserina.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Lasiosphaeriales; Podospora.

NCBI_TaxID=5145;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=96009891; PubMed=7557402;
 RA Saude S., Turco B., Bequeret J.;
 RT "A gene responsible for vegetative incompatibility in the fungus
 RT podospira anserina encodes a protein with a GTP-binding motif and G
 RT beta homologous domain."
 RL Gene 162:135-139(1995).
 CC -1- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
 CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
 CC HET-C.
 CC -1- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: CONTAINS 1 NACHT DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L28125; AA85775.1; -
 CC DR HSSP: P04901; 1786;
 CC DR InterPro: IPR001680; WD40.
 CC DR Pfam: PF00400; WD40; 10.
 CC DR PRINTS: PR00320; GPROTEINBRPT.
 CC DR SMART: SM00320; WD40; 10.
 CC DR PROSITE: PS50837; NACHT; 1.
 CC DR PROSITE: PS00678; WD_REPEATS_1; 10.
 CC DR PROSITE: PS50082; WD_REPEATS_2; 10.
 CC DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 CC DR GTP-binding; Repeat; WD repeat
 CC KW DOMAIN 294 629 NACHT
 CC FT NP_BIND 300 307 GTP (POTENTIAL).
 CC FT REPEAT 839 869 WD 1.
 CC FT REPEAT 881 911 WD 2.
 CC FT REPEAT 923 953 WD 3.
 CC FT REPEAT 965 995 WD 4.
 CC FT REPEAT 1007 1037 WD 5.
 CC FT REPEAT 1049 1079 WD 6.
 CC FT REPEAT 1091 1121 WD 7.
 CC FT REPEAT 1133 1163 WD 8.
 CC FT REPEAT 1175 1205 WD 9.
 CC FT REPEAT 1217 1247 WD 10.
 CC FT REPEAT 1259 1289 WD 11.
 CC FT REPEAT 1301 1331 WD 12.
 CC FT REPEAT 1356 1386 WD 13.
 CC FT REPEAT 1408 1438 WD 14.
 CC FT REPEAT 1460 1490 WD 15.
 CC FT REPEAT 1512 1542 WD 16.
 CC FT REPEAT 1564 1594 WD 17.
 CC FT REPEAT 1616 1646 WD 18.
 CC FT REPEAT 1668 1698 WD 19.
 CC FT REPEAT 1720 1750 WD 20.
 CC FT REPEAT 1772 1802 WD 21.
 CC FT REPEAT 1824 1854 WD 22.
 CC FT REPEAT 1876 1906 WD 23.
 CC FT REPEAT 1928 1958 WD 24.
 CC FT REPEAT 1980 2010 WD 25.
 CC FT REPEAT 2032 2062 WD 26.
 CC FT REPEAT 2084 2114 WD 27.
 CC FT REPEAT 2136 2166 WD 28.
 CC FT REPEAT 2188 2218 WD 29.
 CC FT REPEAT 2240 2270 WD 30.
 CC FT REPEAT 2292 2322 WD 31.
 CC FT REPEAT 2344 2374 WD 32.
 CC FT REPEAT 2396 2426 WD 33.
 CC FT REPEAT 2448 2478 WD 34.
 CC FT REPEAT 2500 2530 WD 35.
 CC FT REPEAT 2552 2582 WD 36.
 CC FT REPEAT 2604 2634 WD 37.
 CC FT REPEAT 2656 2686 WD 38.
 CC FT REPEAT 2708 2738 WD 39.
 CC FT REPEAT 2760 2790 WD 40.
 CC FT REPEAT 2812 2842 WD 41.
 CC FT REPEAT 2864 2894 WD 42.
 CC FT REPEAT 2916 2946 WD 43.
 CC FT REPEAT 2968 2998 WD 44.
 CC FT REPEAT 3020 3050 WD 45.
 CC FT REPEAT 3072 3102 WD 46.
 CC FT REPEAT 3124 3154 WD 47.
 CC FT REPEAT 3176 3206 WD 48.
 CC FT REPEAT 3228 3258 WD 49.
 CC FT REPEAT 3280 3310 WD 50.
 CC FT REPEAT 3332 3362 WD 51.
 CC FT REPEAT 3384 3414 WD 52.
 CC FT REPEAT 3436 3466 WD 53.
 CC FT REPEAT 3488 3518 WD 54.
 CC FT REPEAT 3540 3570 WD 55.
 CC FT REPEAT 3592 3622 WD 56.
 CC FT REPEAT 3644 3674 WD 57.
 CC FT REPEAT 3696 3726 WD 58.
 CC FT REPEAT 3748 3778 WD 59.
 CC FT REPEAT 3800 3830 WD 60.
 CC FT REPEAT 3852 3882 WD 61.
 CC FT REPEAT 3904 3934 WD 62.
 CC FT REPEAT 3956 3986 WD 63.
 CC FT REPEAT 4008 4038 WD 64.
 CC FT REPEAT 4060 4090 WD 65.
 CC FT REPEAT 4112 4142 WD 66.
 CC FT REPEAT 4164 4194 WD 67.
 CC FT REPEAT 4216 4246 WD 68.
 CC FT REPEAT 4268 4298 WD 69.
 CC FT REPEAT 4320 4350 WD 70.
 CC FT REPEAT 4372 4402 WD 71.
 CC FT REPEAT 4424 4454 WD 72.
 CC FT REPEAT 4476 4506 WD 73.
 CC FT REPEAT 4528 4558 WD 74.
 CC FT REPEAT 4580 4610 WD 75.
 CC FT REPEAT 4632 4662 WD 76.
 CC FT REPEAT 4684 4714 WD 77.
 CC FT REPEAT 4736 4766 WD 78.
 CC FT REPEAT 4788 4818 WD 79.
 CC FT REPEAT 4840 4870 WD 80.
 CC FT REPEAT 4892 4922 WD 81.
 CC FT REPEAT 4944 4974 WD 82.
 CC FT REPEAT 4996 5026 WD 83.
 CC FT REPEAT 5048 5078 WD 84.
 CC FT REPEAT 5100 5130 WD 85.
 CC FT REPEAT 5152 5182 WD 86.
 CC FT REPEAT 5204 5234 WD 87.
 CC FT REPEAT 5256 5286 WD 88.
 CC FT REPEAT 5308 5338 WD 89.
 CC FT REPEAT 5360 5390 WD 90.
 CC FT REPEAT 5412 5442 WD 91.
 CC FT REPEAT 5464 5494 WD 92.
 CC FT REPEAT 5516 5546 WD 93.
 CC FT REPEAT 5568 5598 WD 94.
 CC FT REPEAT 5620 5650 WD 95.
 CC FT REPEAT 5672 5702 WD 96.
 CC FT REPEAT 5724 5754 WD 97.
 CC FT REPEAT 5776 5806 WD 98.
 CC FT REPEAT 5828 5858 WD 99.
 CC FT REPEAT 5880 5910 WD 100.
 CC FT REPEAT 5932 5962 WD 101.
 CC FT REPEAT 5984 6014 WD 102.
 CC FT REPEAT 6036 6066 WD 103.
 CC FT REPEAT 6088 6118 WD 104.
 CC FT REPEAT 6140 6170 WD 105.
 CC FT REPEAT 6192 6222 WD 106.
 CC FT REPEAT 6244 6274 WD 107.
 CC FT REPEAT 6296 6326 WD 108.
 CC FT REPEAT 6348 6378 WD 109.
 CC FT REPEAT 6400 6430 WD 110.
 CC FT REPEAT 6452 6482 WD 111.
 CC FT REPEAT 6504 6534 WD 112.
 CC FT REPEAT 6556 6586 WD 113.
 CC FT REPEAT 6608 6638 WD 114.
 CC FT REPEAT 6660 6690 WD 115.
 CC FT REPEAT 6712 6742 WD 116.
 CC FT REPEAT 6764 6794 WD 117.
 CC FT REPEAT 6816 6846 WD 118.
 CC FT REPEAT 6868 6898 WD 119.
 CC FT REPEAT 6920 6950 WD 120.
 CC FT REPEAT 6972 7002 WD 121.
 CC FT REPEAT 7024 7054 WD 122.
 CC FT REPEAT 7076 7106 WD 123.
 CC FT REPEAT 7128 7158 WD 124.
 CC FT REPEAT 7180 7210 WD 125.
 CC FT REPEAT 7232 7262 WD 126.
 CC FT REPEAT 7284 7314 WD 127.
 CC FT REPEAT 7336 7366 WD 128.
 CC FT REPEAT 7388 7418 WD 129.
 CC FT REPEAT 7440 7470 WD 130.
 CC FT REPEAT 7492 7522 WD 131.
 CC FT REPEAT 7544 7574 WD 132.
 CC FT REPEAT 7596 7626 WD 133.
 CC FT REPEAT 7648 7678 WD 134.
 CC FT REPEAT 7700 7730 WD 135.
 CC FT REPEAT 7752 7782 WD 136.
 CC FT REPEAT 7804 7834 WD 137.
 CC FT REPEAT 7856 7886 WD 138.
 CC FT REPEAT 7908 7938 WD 139.
 CC FT REPEAT 7960 7990 WD 140.
 CC FT REPEAT 8012 8042 WD 141.
 CC FT REPEAT 8064 8094 WD 142.
 CC FT REPEAT 8116 8146 WD 143.
 CC FT REPEAT 8168 8198 WD 144.
 CC FT REPEAT 8220 8250 WD 145.
 CC FT REPEAT 8272 8302 WD 146.
 CC FT REPEAT 8324 8354 WD 147.
 CC FT REPEAT 8376 8406 WD 148.
 CC FT REPEAT 8428 8458 WD 149.
 CC FT REPEAT 8480 8510 WD 150.
 CC FT REPEAT 8532 8562 WD 151.
 CC FT REPEAT 8584 8614 WD 152.
 CC FT REPEAT 8636 8666 WD 153.
 CC FT REPEAT 8688 8718 WD 154.
 CC FT REPEAT 8740 8770 WD 155.
 CC FT REPEAT 8792 8822 WD 156.
 CC FT REPEAT 8844 8874 WD 157.
 CC FT REPEAT 8896 8926 WD 158.
 CC FT REPEAT 8948 8978 WD 159.
 CC FT REPEAT 9000 9030 WD 160.
 CC FT REPEAT 9052 9082 WD 161.
 CC FT REPEAT 9104 9134 WD 162.
 CC FT REPEAT 9156 9186 WD 163.
 CC FT REPEAT 9208 9238 WD 164.
 CC FT REPEAT 9260 9290 WD 165.
 CC FT REPEAT 9312 9342 WD 166.
 CC FT REPEAT 9364 9394 WD 167.
 CC FT REPEAT 9416 9446 WD 168.
 CC FT REPEAT 9468 9498 WD 169.
 CC FT REPEAT 9520 9550 WD 170.
 CC FT REPEAT 9572 9602 WD 171.
 CC FT REPEAT 9624 9654 WD 172.
 CC FT REPEAT 9676 9706 WD 173.
 CC FT REPEAT 9728 9758 WD 174.
 CC FT REPEAT 9780 9810 WD 175.
 CC FT REPEAT 9832 9862 WD 176.
 CC FT REPEAT 9884 9914 WD 177.
 CC FT REPEAT 9936 9966 WD 178.
 CC FT REPEAT 9988 10018 WD 179.
 CC FT REPEAT 10040 10070 WD 180.
 CC FT REPEAT 10092 10122 WD 181.
 CC FT REPEAT 10144 10174 WD 182.
 CC FT REPEAT 10196 10226 WD 183.
 CC FT REPEAT 10248 10278 WD 184.
 CC FT REPEAT 10300 10330 WD 185.
 CC FT REPEAT 10352 10382 WD 186.
 CC FT REPEAT 10404 10434 WD 187.
 CC FT REPEAT 10456 10486 WD 188.
 CC FT REPEAT 10508 10538 WD 189.
 CC FT REPEAT 10560 10590 WD 190.
 CC FT REPEAT 10612 10642 WD 191.
 CC FT REPEAT 10664 10694 WD 192.
 CC FT REPEAT 10716 10746 WD 193.
 CC FT REPEAT 10768 10798 WD 194.
 CC FT REPEAT 10820 10850 WD 195.
 CC FT REPEAT 10872 10902 WD 196.
 CC FT REPEAT 10924 10954 WD 197.
 CC FT REPEAT 10976 11006 WD 198.
 CC FT REPEAT 11028 11058 WD 199.
 CC FT REPEAT 11080 11110 WD 200.
 CC FT REPEAT 11132 11162 WD 201.
 CC FT REPEAT 11184 11214 WD 202.
 CC FT REPEAT 11236 11266 WD 203.
 CC FT REPEAT 11288 11318 WD 204.
 CC FT REPEAT 11340 11370 WD 205.
 CC FT REPEAT 11392 11422 WD 206.
 CC FT REPEAT 11444 11474 WD 207.
 CC FT REPEAT 11496 11526 WD 208.
 CC FT REPEAT 11548 11578 WD 209.
 CC FT REPEAT 11600 11630 WD 210.
 CC FT REPEAT 11652 11682 WD 211.
 CC FT REPEAT 11704 11734 WD 212.
 CC FT REPEAT 11756 11786 WD 213.
 CC FT REPEAT 11808 11838 WD 214.
 CC FT REPEAT 11860 11890 WD 215.
 CC FT REPEAT 11912 11942 WD 216.
 CC FT REPEAT 11964 11994 WD 217.
 CC FT REPEAT 12016 12046 WD 218.
 CC FT REPEAT 12068 12098 WD 219.
 CC FT REPEAT 12120 12150 WD 220.
 CC FT REPEAT 12172 12202 WD 221.
 CC FT REPEAT 12224 12254 WD 222.
 CC FT REPEAT 12276 12306 WD 223.
 CC FT REPEAT 12328 12358 WD 224.
 CC FT REPEAT 12380 12410 WD 225.
 CC FT REPEAT 12432 12462 WD 226.
 CC FT REPEAT 12484 12514 WD 227.
 CC FT REPEAT 12536 12566 WD 228.
 CC FT REPEAT 12588 12618 WD 229.
 CC FT REPEAT 12640 12670 WD 230.
 CC FT REPEAT 12692 12722 WD 231.
 CC FT REPEAT 12744 12774 WD 232.
 CC FT REPEAT 12796 12826 WD 233.
 CC FT REPEAT 12848 12878 WD 234.
 CC FT REPEAT 12900 12930 WD 235.
 CC FT REPEAT 12952 12982 WD 236.
 CC FT REPEAT 13004 13034 WD 237.
 CC FT REPEAT 13056 13086 WD 238.
 CC FT REPEAT 13108 13138 WD 239.
 CC FT REPEAT 13160 13190 WD 240.
 CC FT REPEAT 13212 13242 WD 241.
 CC FT REPEAT 13264 13294 WD 242.
 CC FT REPEAT 13316 13346 WD 243.
 CC FT REPEAT 13368 13398 WD 244.
 CC FT REPEAT 13420 13450 WD 245.
 CC FT REPEAT 13472 13502 WD 246.
 CC FT REPEAT 13524 13554 WD 247.
 CC FT REPEAT 13576 13606 WD 248.
 CC FT REPEAT 13628 13658 WD 249.
 CC FT REPEAT 13680 13710 WD 250.
 CC FT REPEAT 13732 13762 WD 251.
 CC FT REPEAT 13784 13814 WD 252.
 CC FT REPEAT 13836 13866 WD 253.
 CC FT REPEAT 13888 13918 WD 254.
 CC FT REPEAT 13940 13970 WD 255.
 CC FT REPEAT 13992 14022 WD 256.
 CC FT REPEAT 14044 14074 WD 257.
 CC FT REPEAT 14096 14126 WD 258.
 CC FT REPEAT 14148 14178 WD 259.
 CC FT REPEAT 14200 14230 WD 260.
 CC FT REPEAT 14252 14282 WD 261.
 CC FT REPEAT 14304 14334 WD 262.
 CC FT REPEAT 14356 14386 WD 263.
 CC FT REPEAT 14408 14438 WD 264.
 CC FT REPEAT 14460 14490 WD 265.
 CC FT REPEAT 14512 14542 WD 266.
 CC FT REPEAT 14564 14594 WD 267.
 CC FT REPEAT 14616 14646 WD 268.
 CC FT REPEAT 14668 14698 WD 269.
 CC FT REPEAT 14720 14750 WD 270.
 CC FT REPEAT 14772 14802 WD 271.
 CC FT REPEAT 14824 14854 WD 272.
 CC FT REPEAT 14876 14906 WD 273.
 CC FT REPEAT 14928 14958 WD 274.
 CC FT REPEAT 14980 15010 WD 275.
 CC FT REPEAT 15032 15062 WD 276.
 CC FT REPEAT 15084 15114 WD 277.
 CC FT REPEAT 15136 15166 WD 278.
 CC FT REPEAT 15188 15218 WD 279.
 CC FT REPEAT 15240 15270 WD 280.
 CC FT REPEAT 15292 15322 WD 281.
 CC FT REPEAT 15344 15374 WD 282.
 CC FT REPEAT 15396 15426 WD 283.
 CC FT REPEAT 15448 15478 WD 284.
 CC FT REPEAT 15500 15530 WD 285.
 CC FT REPEAT 15552 15582 WD 286.
 CC FT REPEAT 15604 15634 WD 287.
 CC FT REPEAT 15656 15686 WD 288.
 CC FT REPEAT 15708 15738 WD 289.
 CC FT REPEAT 15760 15790 WD 290.
 CC FT REPEAT 15812 15842 WD 291.
 CC FT REPEAT 15864 15894 WD 292.
 CC FT REPEAT 15916 15946 WD 293.
 CC FT REPEAT 15968 15998 WD 294.
 CC FT REPEAT 16020 16050 WD 295.
 CC FT REPEAT 16072 16102 WD 296.
 CC FT REPEAT 16124 16154 WD 297.
 CC FT REPEAT 16176 16206 WD 298.
 CC FT REPEAT 16228 16258 WD 299.
 CC FT REPEAT 16280 16310 WD 300.
 CC FT REPEAT 16332 16362 WD 301.
 CC FT REPEAT 16384 16414 WD 302.
 CC FT REPEAT 16436 16466 WD 303.
 CC FT REPEAT 16488 16518 WD 304.
 CC FT REPEAT 16540 16570 WD 305.
 CC FT REPEAT 16592 16622 WD 306.
 CC FT REPEAT 16644 16674 WD 307.
 CC FT REPEAT 16696 16726 WD 308.
 CC FT REPEAT 16748 16778 WD 309.
 CC FT REPEAT 16800 16830 WD 310.
 CC FT REPEAT 16852 16882 WD 311.
 CC FT REPEAT 16904 16934 WD 312.
 CC FT REPEAT 16956 16986 WD 313.
 CC FT REPEAT 17008 17038 WD 314.
 CC FT REPEAT 17060 17090 WD 315.
 CC FT REPEAT 17112 17142 WD 316.
 CC FT REPEAT 17164 17194 WD 317.
 CC FT REPEAT 17216 17246 WD 318.
 CC FT REPEAT 17268 17298 WD 319.
 CC FT REPEAT 17320 17350 WD 320.
 CC FT REPEAT 17372 17402 WD 321.
 CC FT REPEAT 17424 17454 WD 322.
 CC FT REPEAT 17476 17506 WD 323.
 CC FT REPEAT 17528 17558 WD 324.
 CC FT REPEAT 17580 17610 WD 325.
 CC FT REPEAT 17632 17662 WD 326.
 CC FT REPEAT 17684 17714 WD 327.
 CC FT REPEAT 17736 17766 WD 328.
 CC FT REPEAT 17788 17818 WD 329.
 CC FT REPEAT 17840 17870 WD 330.
 CC FT REPEAT 17892 17922 WD 331.
 CC FT REPEAT 17944 17974 WD 332.
 CC FT REPEAT 17996 18026 WD 333.
 CC FT REPEAT 18048 18078 WD 334.
 CC FT REPEAT 18100 18130 WD 335.
 CC FT REPEAT 18152 18182 WD 336.
 CC FT REPEAT 18204 18234 WD 337.
 CC FT REPEAT 18256 18286 WD 338.
 CC FT REPEAT 18308 18338 WD 339.
 CC FT REPEAT 18360 18390 WD 340.
 CC FT REPEAT 18412 18442 WD 341.
 CC FT REPEAT 18464 18494 WD 342.
 CC FT REPEAT 18516 18546 WD 343.
 CC FT REPEAT 18568 18598 WD 344.
 CC FT REPEAT 18620 18650 WD 345.
 CC FT REPEAT 18672 18702 WD 346.
 CC FT REPEAT 18724 18754 WD 347.
 CC FT REPEAT 18776 18806 WD 348.
 CC FT REPEAT 18828 18858 WD 349.
 CC FT REPEAT 18880 18910 WD 350.
 CC FT REPEAT 18932 18962 WD 351.
 CC FT REPEAT 18984 19014 WD 352.
 CC FT REPEAT 19036 19066 WD 353.
 CC FT REPEAT 19088 19118 WD 354.
 CC FT REPEAT 19140 19170 WD 355.
 CC FT REPEAT 19192 19222 WD 356.
 CC FT REPEAT 19244 19274 WD 357.
 CC FT REPEAT 19296 19326 WD 358.
 CC FT REPEAT 19348 19378 WD 359.
 CC FT REPEAT 19400 19430 WD 360.
 CC FT REPEAT 19452 19482 WD 361.
 CC FT REPEAT 19504 19534 WD 362.
 CC FT REPEAT 19556 19586 WD 363.
 CC FT REPEAT 19608 19638 WD 364.
 CC FT REPEAT 19660 19690 WD 365.
 CC FT REPEAT 19712 19742 WD 366.
 CC FT REPEAT 19764 19794 WD 367.
 CC FT REPEAT 19816 19846 WD 368.
 CC FT REPEAT 19868 19898 WD 369.
 CC FT REPEAT 19920 19950 WD 370.
 CC FT REPEAT 19972 20002 WD 371.
 CC FT REPEAT 20024 20054 WD 372.
 CC FT REPEAT 20076 20106 WD 373.
 CC FT REPEAT 20128 20158 WD 374.
 CC FT REPEAT 20180 20210 WD 375.
 CC FT REPEAT 20232 20262 WD 376.
 CC FT REPEAT 20284 20314 WD 377.
 CC FT REPEAT 20336 20366 WD 378.
 CC FT REPEAT 20388 20418 WD 379.
 CC FT REPEAT 20440 20470 WD 380.
 CC FT REPEAT 20492 20522 WD 381.
 CC FT REPEAT 20544 20574 WD 382.
 CC FT REPEAT 20596 20626 WD 383.
 CC FT REPEAT 20648 20678 WD 384.
 CC FT REPEAT 20700 20730 WD 385.
 CC FT REPEAT 20752 20782 WD 386.
 CC FT REPEAT 20804 20834 WD 387.
 CC FT REPEAT 20856 20886 WD 388.
 CC FT REPEAT 20908 20938 WD 389.
 CC FT REPEAT 20960 20990 WD 390.
 CC FT REPEAT 21012 21042 WD 391.
 CC FT REPEAT 21064 21094 WD 392.
 CC FT REPEAT 21116 21146 WD 393.
 CC FT REPEAT 21168 21198 WD 394.
 CC FT REPEAT 21220 21250 WD 395.
 CC FT REPEAT 21272 21302 WD 396.
 CC FT REPEAT 21324 21354 WD 397.
 CC FT REPEAT 21376 21406 WD 398.
 CC FT REPEAT 21428 21458 WD 399.
 CC FT REPEAT 21480 21510 WD 400.
 CC FT REPEAT 21532 21562 WD 401.
 CC FT REPEAT 21584 21614 WD 402.
 CC FT REPEAT 21636 21666 WD 403.
 CC FT REPEAT 21688 21718 WD 404.
 CC FT REPEAT 21740 21770 WD 405.
 CC FT REPEAT 21792 21822 WD 406.
 CC FT REPEAT 21844 21874 WD 407.
 CC FT REPEAT 21896 21926 WD 408.
 CC FT REPEAT 21948 21978 WD 409.
 CC FT REPEAT 22000 22030 WD 410.
 CC FT REPEAT 22052 22082 WD 411.
 CC FT REPEAT 22104 22134 WD 412.
 CC FT REPEAT 22156 22186 WD 413.
 CC FT REPEAT 22208 22238 WD 414.
 CC FT REPEAT 22260 22290 WD 415.
 CC FT REPEAT 22312 22342 WD 416.
 CC FT REPEAT 22364 22394 WD 417.
 CC FT REPEAT 22416 22446 WD 418.
 CC FT REPEAT 22468 22498 WD 419.
 CC FT REPEAT 22520 22550 WD 420.
 CC FT REPEAT 22572 22602 WD 421.
 CC FT REPEAT 22624 22654 WD 422.
 CC FT REPEAT 22676 22706 WD 423.
 CC FT REPEAT 22728 22758 WD 424.
 CC FT REPEAT 22780 22810 WD 425.
 CC FT REPEAT 22832 22862 WD 426.
 CC FT REPEAT 22884 22914 WD 427.
 CC FT REPEAT 22936 22966 WD 428.
 CC FT REPEAT 22988 23018 WD 429.
 CC FT REPEAT 23040 23070 WD 430.
 CC FT REPEAT 23092 23122 WD 431.
 CC FT REPEAT 23144 23174 WD 432.
 CC FT REPEAT 23196 23226 WD 433.
 CC FT REPEAT 23248 23278 WD 434.
 CC FT REPEAT 23300 23330 WD 435.
 CC FT REPEAT 23352 23382 WD 436.
 CC FT REPEAT 23404 23434 WD 437.
 CC FT REPEAT 23456 23486 WD 438.
 CC FT REPEAT 23508 23538 WD 439.
 CC FT REPEAT 23560 23590 WD 440.
 CC FT REPEAT 23612 23642 WD 441.
 CC FT REPEAT 23664 23694 WD 442.
 CC FT REPEAT 23716 23746 WD 443.
 CC FT REPEAT 23768 23798 WD 444.
 CC FT REPEAT 23820 23850 WD 445.
 CC FT REPEAT 23872 23902 WD 446.
 CC FT REPEAT 23924 23954 WD 447.
 CC FT REPEAT 23976 24006 WD 448.
 CC FT REPEAT 24028 24058 WD 449.
 CC FT REPEAT 24080 24110 WD 450.
 CC FT REPEAT 24132 24162 WD 451.
 CC FT REPEAT 24184 24214 WD 452.
 CC FT REPEAT 24236 24266 WD 453.
 CC FT REPEAT 24288 24318 WD 454.
 CC FT REPEAT 24340 24370 WD 455.
 CC FT REPEAT 24392 24422 WD 456.
 CC FT REPEAT 24444 24474 WD 457.
 CC FT REPEAT 24496 24526 WD 458.
 CC FT REPEAT 24548 24578 WD 459.
 CC FT REPEAT 24600 24630 WD 460.

Db 265 -----GEYQREGIAPVSD-----TEARRTTSQSWYVYINPCKRV 301
 QY 169 FYLNLVWSSHNVLAVGLGNVYLMNACSSKVTKLCDGVDDVCVGAAGCTHLAVCT 228
 Db 302 FMINVHTLH-----PSYCCVCKSSNNGKTLATGC 332
 QY 229 NNGVQVWDAAKCKRISME-----GHLRVGALAMS--SSLSGGGDKNIYQDRT 280
 Db 333 NQA-ANVEDVOTGKLTFTLHESPPDSRDLVYRTAFSDGKYVLATGEDRQIKMLDST 391
 QY 281 QE-DEVSRLSGHSEVCGKMSVDNDELASGDNKLVYMNQHSOTPYLKYCEHTAIVKA 339
 Db 392 QKVXYV--FSGHEODIYSLDFSHNGRFYVSGSDRTARLMPVERGOCILK-LEIENGYTA 448
 QY 340 IAMSPLHGLASGGGTADRCIRFNNVTNNLSLSCMDTGSQVCMVMSKNNELVSTHY 399
 Db 449 IALSPMDQFLAV--GSLDOIIRVMSVSGTLVERLEGHKESVYIANSPPSILL--GS 503
 QY 400 SONQIIVWRYPTMSKL-----ATLGHTRVLYALISPDGQTLVTCAGDETLR 447
 Db 504 LDKTIKVMELQATRSVGLSAIKPECICAKATYTGHTDFLVAVSPDSRMGLSGSKDSMO 563
 QY 448 FNNV 451
 Db 564 FMDL 567

RESULT 8

T012_SCHPO STANDARD; PRT; 586 AA.
 AC 09UG8: 014432: (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional repressor tup12.
 GN TUP12 OR TUP1 OR SPAC630.14C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC NCBI_TaxId=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA McDougall R.C., Rajandream M.A., Barrell B.G., Zimmermann W.,
 RA Wambolt R.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 26-586 FROM N.A.
 RC STRAIN=972;
 RA Braun B.R.;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 RP CHARACTERIZATION.
 RC MEDLINE=20036803; PubMed=10567571;
 RA Mukai Y., Matsuo E., Roth S.Y., Harashima S.;
 RT "Conservation of histone binding and transcriptional repressor
 RL Mol. Cell. Biol. 19:8461-8468(1999)."
 CC - FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC - SIMILARITY: BELONGS TO THE TUP1 FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC entitled and this statement is not removed. Usage by and for commercial
 CC or send an email to license@ebi-sib.ch).
 DR EMBL: AL109832; CAB52736.1;
 DR EMBL: U92792; AAB81475.2;

DR InterPro: IPR001632; Gprotein_B.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PRO0319; GPROTEINB.
 DR PRINTS: PRO0320; GPROTEINBPT.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; 3.
 DR PROSITE: PS50082; WD_REPEATS_2; 6.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repressor; Repeat; WD repeat.
 FT REPEAT 332 325 WD 1.
 FT REPEAT 374 371 WD 2.
 FT REPEAT 374 413 WD 3.
 FT REPEAT 415 454 WD 4.
 FT REPEAT 456 495 WD 5.
 FT REPEAT 510 549 WD 6.
 FT REPEAT 552 585 WD 7.
 SO SEQUENCE 586 AA; 64192 MW; 79C991ABDDC7D095 CRC64;

Query Match
 Best Local Similarity 9.8%; Score 250; DB 1; Length 586;
 Matches 117; Conservative 73; Mismatches 175; Indels 146; Gaps 23;

QY 5 GNRNPPPTSTVRDNDSPPP-----EPSESLR-----HVSRY-INSNH 41
 Db 130 GNAGPFNS-----DNSKPLILNNGSSGCTFKNLNSPAIDSDGTVLAPIQTSNVNDSQYY 185
 QY 42 TSP-----SRITYSD--RFIPSNASAKFALFDINPTREGDDSSAVYTLRLALFGFD 93
 Db 186 SSPHRAVAGATMAGSAMRTFSPN-----LPLGHPPPPSDSANSVPPIAFLVNGK 238
 QY 94 VAG-VPTPEKTDSPSMTLPNNINIRYKTETRQSMHSLSPMDDEYVGNHSPVAPARKV 152
 Db 239 VSGEPYPAEI-IFTSNVNRE-----EKDMTVYNSN-----VPKKE 273
 QY 153 PRSPKVLVDAPODDDYLLVDMSSH-NVLAVGLGNCVYLMNACSSKVTKLC----- 204
 Db 274 P-PLVQLHTLEHTSVICVYRFESADGKFLATGCRAMAVENVETGKLIITLLODESSKR 331
 QY 205 --DLGVDCVCSVGMARGTALAVGTNNGKVOIMPAARCKIRSEGHRLVGLAMSSS 262
 Db 332 EGDL---YVRSVAFSPDGKYLATGVEDQIRINDIAO----- 365
 QY 263 LLSGGGDKNIYQDRTIREDVSKLSGHKSEVCGKMSYDNRELASGDNKLVYMNQ 322
 Db 366 -----KRYR-----LLTGHDEQIYSLDFSKDKGLVSGSDRTVCLMDVE 406
 QY 323 STQPVLYKEHT-AAVKALAMSPHLHGLASGGGTADRCIRFNNVTNNLSLSCMDTGSQV 381
 Db 407 AGEQ--KLILHTDDGVTVVMSPDQFLA--GSLDKVIRIWTSSGTLVEQLHGEESV 461
 QY 382 CNLWMSKNNELVSHGSSQNOIIVWRYPTMSKLA-----TYLGHYRVLYLA 429
 Db 462 YSVASPPQKIVS--GSLDNTIKLMELQCVSNVASMKEGICKQTFTGHKDPILSVT 519
 QY 430 ISPDQTLVTAAGDETLRFNVWVFPSPKQNT 460
 Db 520 VSPDGKWLISGSKDRITQFWS--DSPHSQLT 549
 RESULT 9
 YCW2_YEAST STANDARD; PRT; 515 AA.
 AC P23382;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypochemical 57.0 kDa Trp-asp repeats containing protein in CPR4-SSK22
 DE Interpeptide region.
 GN YCR072C OR YCR72C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Wed May 22 11:09:50 2002

us-09-701-572-2.rsp

Page 7

Saccharomycetales: Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RA Ballester J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RL [2]
RA REVISIONS.
RA Jimenez A.,
RA Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RL -1 SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X59720; CAA42270.1;
CC PIR: S19487; S19487.
CC SCD: S000668; YCR072C.
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 8.
CC PRINTS: PR00320; GPROTEINBPT.
CC SMART: SM00320; WD40; 8.
CC PROSITE: PS00678; WD_REPEATS_1; 5.
CC PROSITE: PS50082; WD_REPEATS_2; 7.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC DR Hypothetical protein; Repeat; WD repeat.
CC KW Repeat; WD 1.
CC FT Repeat; 141 172 WD 2.
CC FT Repeat; 184 214 WD 2.
CC FT Repeat; 227 264 WD 3.
CC FT Repeat; 276 305 WD 4.
CC FT Repeat; 318 387 WD 5.
CC FT Repeat; 400 430 WD 6.
CC FT Repeat; 442 472 WD 7.
CC FT Repeat; 484 514 WD 8.
CC FT Repeat; 515 AA; 57040 MW; DFE603B79BFB530A CRC64;
CC SEQUENCE
SO
Query Match 9.3%; Score 237; DB 1; Length 515;
Best Local Similarity 20.7%; Pred. No. 3.5e-11;
Matches 104; Conservative 74; Mismatches 208; Indels 116; Gaps 18;
24 PSEPLRHVSRMINSNHYTSPRTYS-DRFIPRSASKFFALFDINTPTGRRDSSAYT 82
52 PAISEKOLEELNOLNNTSDPVPYFSCIOGKRAS-----DPKTIIDITDNLYS 103
83 TLLRTALFGPDVAGPVEPEKTPSPMLPRNIFRYKTETRQS-----MHSISP 131
104 SLIK-----PGYNSTEDQITLLTPRAVKKVPRYSSSAINGSTTLCASAP 153
132 FMDDFEVPGVNHSPPV-----APRYPRSPYKVDAPALODEFLNLVWS-SHNVLA 183
154 HTSSRAVATGAGDNTARIMDCDTOTPMHTLKGHY-----WVICVNSPDEVI 202
184 VG-IGNCVYLAMNCSKATKICLDLGVDCVCSGVN-----AQRT--HLAVGTNNCKVOI 235
203 TGSMDNTIRLMDPKSGQGLDGLRGHSKWITSLSEWPIHLVKKPSKPLASSKIDGITIKI 262
226 WDAARCKIRSMGCHRLVGLALW-SSSILSSGGRDKNIYORDIRTOEDFYSKLSGKSE 294
263 WDTVRVCOYTMGHTNSVSCVWGGGGLYSCHDRYRVWDINSGRCINILKSHAHW 322
295 VCGIKMSVD-----NRE--LASGNDNK 315
323 VMLSLSTDYALRIGAFDHTGKKRSTPEAKALENEYKICKKNNGSEMMVATSDYT 382
316 LFWVND-HSTQVPLKYCHETAAVKAIAIAMSPLHGLLAGGGTADRCIRWN-----TT 367

DB 383 MFLNPLKSTKPIAMTCHOKLVNNAFSPDGRIVSA---SFDNSIKIMDRDCKFI 439
OY 368 TNSHSCMPTQSGVQCNLVNWSKNVLYSTHGYSONOITVRYPTMSKLTATLGHYRYLY 427
DB 440 FRGHT-----ASVQVAMSSDCRLVSCS--KOTITLKVMDVFRKLSDLPKHDEVYT 491
OY 428 LAISPDGQITVTCAGDETLRFN 449
DB 492 VDMSSVGRKVCSSGDKRMYRLW 513

RESULT 10
ID TUP1_CANAL STANDARD; PRT: 514 AA.
AC P56093;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional repressor TUP1.
GN TUP1 OR CHP1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5476;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MO-1;
RC MEDLINE=97349341; PubMed=9204892;
RX Braun B.R., Johnson A.D.,
RT "Control of filament formation in Candida albicans by the
RT transcriptional repressor TUP1."
RL Science 277:105-109(1997).
CC -1- FUNCTION: REPRESENTS TRANSCRIPTION BY RNA POLYMERASE II. REPRESENTS
CC -1- GENES RESPONSIBLE FOR INITIATING FILAMENTOUS GROWTH AND THIS
CC REPRESSION IS LIFTED UNDER INDUCING ENVIRONMENTAL CONDITIONS.
CC
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF005741; AAB63195.1;
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 7.
CC PRINTS: PR00320; GPROTEINBPT.
CC SMART: SM00320; WD40; 7.
CC PROSITE: PS00678; WD_REPEATS_1; 3.
CC PROSITE: PS50082; WD_REPEATS_2; 6.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC DR Transcription regulation; Repressor; Repeat; WD repeat.
CC KW Repeat; 200 240 WD 1.
CC FT Repeat; 255 296 WD 2.
CC FT Repeat; 297 336 WD 3.
CC FT Repeat; 338 377 WD 4.
CC FT Repeat; 387 426 WD 5.
CC FT Repeat; 434 473 WD 6.
CC FT Repeat; 476 513 WD 7.
CC FT Repeat; 514 516 POLY-GLN.
CC FT DOMAIN 97 116
CC SEQUENCE 514 AA; 57821 MW; 3A88A5413DFB69F9 CRC64;
SO
Query Match 8.6%; Score 221; DB 1; Length 514;
Best Local Similarity 21.8%; Pred. No. 6.3e-10;
Matches 86; Conservative 70; Mismatches 154; Indels 84; Gaps 15;
94 VAGVTEPKTDSPTKMLPRNIFRY---KTETRQSMHSLSPMDDDVPVGVNHSPPVAPR 150
DB 118 VAPPAAPAPPTVSTLSDIKSQIYVNTQANHVKEIPLDLDIAKAN-----PE 171

OY 151 -KVRSPYKVLADAPALODDFYLNLDKSSHNVLAVGLNCVYLMACSSKYTKLDLGYD 209
 DB 172 EKQHLLEYVLYNPAFSKDLIDMVLHSHS----- 202
 OY 210 DVCVSGVMAORGHVLAAGTNGKVQIWDARCKIRSM-----EGRRLRVG 255
 DB 203 SVCCVAFSSDGFATNGCNK-TTQFVNTTGLVAKLIDSSSENKDDNTTASGDLYIR 261
 OY 256 ALAMS--SLLSSGGRDNKIYORDIRTEDEPFYKLSGKSVCGKMSYDNRELASGMD 313
 DB 262 SVCFSPDGLKLAGAEDKLIRIMDLSTR-IKILRGHEDDIYSLDFPDGDRVLSGSGD 320
 OY 314 NKLEFVNNHSTOPLVKYCEHTAAVK--ALAMSPHLGLLASGGGTADRCIRFNNVTN 369
 DB 321 RSVRIWDLRTSQ-----CSLTLSDGCVTVAVSP--DGKLIA-AGSLDRIVRVVDSITG 372
 OY 370 SHLSCMDTSSQ-----VCNLVMSKNVNELVSTHGYSONQIIVRYITYSKL-----A 416
 DB 373 FIVERLDSNGENGNHEDSVYSAFNSNGEQIAS--GSLDRIVKLMHLGKSDKSNCEV 430
 OY 417 TLTGHTYVLYLAISPDGQITVAGDETLRFVN 450
 DB 431 TYIGHKDFVLISVCTPDNEYITLSSGKDRGVTFWD 464

RESULT 11

RCOL_NEUCR
 ID RCOL_NEUCR STANDARD: PRT: 604 AA.
 AC P78706:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Transcriptional repressor rco-1.
 GN RCO-1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariatales; Sordariaceae; Neurospora.
 RN NCBI_TaxID=5141;
 [1]

RA MEDLINE-97042452; PubMed-8887652;
 RA Yamashiro C.T., Ebbole D.J., Lee B.-U., Brown R.E., Bourland C.,
 RT Madl L., Yanofsky C.;
 RT "Characterization of rco-1 of Neurospora crassa, a pleiotropic gene
 affecting growth and development that encodes a homolog of Tup1 of
 Saccharomyces cerevisiae."
 RL Mol. Cell. Biol. 16:6218-6228(1996).
 CC -1- FUNCTION: REPRESSSES TRANSCRIPTION BY RNA POLYMERASE II. MAY BE
 INVOLVED AT SEVERAL STAGES OF CONDITATION AND OTHER GROWTH AND
 DEVELOPMENT PROCESSES. APPEARS TO REGULATE GENES THAT ARE
 EXPRESSED IN ASEYUAL AND SEXUAL SPORE PATHWAYS. MUTANT RCO-1
 ALLELES ARE ACONIDIAL, FEMALE STERILE, HAVE REDUCED GROWTH RATES
 AND FORM HYPAE THAT COIL IN A COUNTERCLOCKWISE DIRECTION,
 OPPOSITE TO THAT OF THE WILD TYPE.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: U57061; AAB37245.1;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PRINTS: PR00320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; 4.
 DR PROSITE: PS50082; WD_REPEATS_2; 6.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repressor; Repeat; WD repeat.

FT REPEAT 295 324 WD 1.
 FT REPEAT 342 372 WD 2.
 FT REPEAT 384 414 WD 3.
 FT REPEAT 425 455 WD 4.
 FT REPEAT 469 499 WD 5.
 FT REPEAT 523 553 WD 6.
 FT REPEAT 565 600 WD 7.
 FT DOMAIN 130 135 POLY-PRO.
 FT DOMAIN 153 160 POLY-PRO.
 FT DOMAIN 244 249 POLY-PRO.
 SQ SEQUENCE 604 AA; 66076 MW; C962B5E558B529A4 CRC64;

Query Match Best Local Similarity 8.5%; Score 217; DB 1; Length 604;
 Matches 109; Conservative 84; Mismatches 202; Indels 136; Gaps 24;

OY 5 GNRNPPPTSTVVRDNP-----PPSPESLRHVSRIKNSNHYT 42
 DB 92 GNMHPPOHNGOQPPAIGLGNVFSALIMAGCGGALVPPPPPOOOPAHNPAPGLQ 151
 OY 43 SP-----SRTIYSDRFIPRSASRFPALPDITPEEGDDSSAYTLLFTALFPGV 94
 DB 152 GPPPPPPPSQPPPOQOYQGGQGNF-----PPQPSYASGPGAKR-GIGRPPA 203
 OY 95 AGVTPPE-----KTDSPM-TLPNNIRYKTETRQSHSLSPMDDPFGVNH 143
 DB 204 GGPATPQITPTIPYNGGASPOVPTHTPTD-----HTRMAOH-----H 242
 OY 144 SPVKAAPKRVPPSPYKVLADAPALODDFYLNLDKSSHNVLAVGLNCVYLMACSSKYTKL 203
 DB 243 QPPPPPS-QTNALSELDPRLPHIKKKMDK-----WVFMAAVPRV 285
 OY 204 CDLGV-----DDCVSGVMAORGHVLAAGTNGKVQIWDARCKIRSMGHR----- 251
 DB 286 LDVELVHTLQHESSVCCVRSMDGKTVAVGCNR-SAQIVDETGEKICLLIDENIDITGD 344
 OY 252 LRVALGAS--SLLSSGGRDNKIYORDIRTEDEPFYKLSGKSVCGKMSYDNRELAS 309
 DB 345 LYRSVCFSPDGLKLAGAEDKLIRIMDLSTR-TIRMTFCHEDDIYSLDFSDGRTIAS 403
 OY 310 GGDNDKLFVNNHSTQ--PVLKCEHTAAVKALAMSPHLGLLASGGGTADRCIRFNNVT 367
 DB 404 GSGDKTVRLMDIETGQNTSVLSIED---GVTVASIDKQFVA--GSLDRSVAVMD-- 455
 OY 368 TNSHLSMDTG-----SOYCNLWMSKNVNELVSTHGYSONQIIVR-----YPMRK-- 414
 DB 456 MGVLAEHLRLEPGDKHDSVYSAFNSNGEQIAS--GSLDKTKMKELSNAPGIPSSAPPK 513
 OY 415 ----LATLTGHTYVLYLAISPDGQITVAGDETLRFVNFPSPKSONTE 461
 DB 514 GGRCTKTFEGHRDFVLSVALTPDSQWVLSGSKDRGVTFWD-----PRTGHTQ 560

RESULT 12

PRLL_ARATH
 ID PRLL_ARATH STANDARD: PRT: 486 AA.
 AC 042384:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE PPL1/PP2A phosphatases pleiotropic regulator PRL1.
 GN PRL1 OR AT4G15900 or PL3990W.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC NCBI_TaxID=3702;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Nemeth K., Putnoy P., Stankovic B., Bako L., Mathur J., Redel G.P.,
 RA Scheil J., Koncz C.;

Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.

RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RA [21]
 RA SEQUENCE FROM N.A.
 RA STRAIN-CY COLUMBIA;
 RA MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Scheller C., Mambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duestenhoef A., Stiekema W., Entlan K.D., Terry N.,
 RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Schmidheini T.,
 RA Kreitz M., Delseny M., Puigdomenech P., Watson M., Schmidt I.,
 RA Reichert B., Portelle D., Perez-Alonso M., Doutry M., Benoit I.,
 RA Vos P., Heideisel J., Zimmermann W., Medler H., Ridley P.,
 RA Landman S.-A., McCullagh B., Bilham L., Robben J., Vandenbusche F.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbusche F.,
 RA Braeken M., Willems I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Willems I., Bohe G., Rampsperger U., Hilbert H., Brun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirse W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetler P.,
 RA Bernier S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H.,
 RA De Keyser A., Buysaert C., Giejen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S., Mayes R.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIlroy S.,
 RA Petter A., Rajandream M.A., Lyne M., Grimm M., Loehner T.-H.,
 RA Bokora D., Bloeker H., Scharfe M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Partmann B., Grandet R., Pivard E.,
 RA Dose S., Argirion A., Vitale D., Liguori R., Fieber R.,
 RA Massenet O., Angillet F., Schmidt W., Lechman A., Aubourg S.,
 RA Schnabl S., Hiller R., Berger C., Monfort A., Casaberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Francis P., Bielke C.,
 RA Reijnen L., Schwarz S., Scholler P., Heber S., Stocker S.,
 RA Fricman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spigel L.,
 RA Sekhon M., Murray J., Sheet P., Cortes M., Abou-Threidh J.,
 RA Lestrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton L., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Keads E., Andrews S., Geisel C., Layman D.,
 RA Nelson J., Speth J., Ryan E., Jones C., Cotton M., Joshi C.,
 RA Du H., Ali J., Bergoff A., Jones C., Sun H., Lamar B., Yordan C.,
 RA Antoniou B., Zidanic M., Strong C., Vill D., Shekher M., Makero A., Shah R.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Makero A., Shah R.,
 RA Swaby I.K., Shohdy N., Hasegawa A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Rodriguez M., Hoffman J., Tili S.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.,
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RA thaliana."
 RA Nature 402:769-777(1999).
 RA FUNCTION: PEPTIDOTIC REGULATOR OF P1 AND P2A PHOSPHATASES.
 CC - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC - SIMILARITY: BELONGS TO THE PRL/PR2 FAMILY OF WD-REPEAT PROTEINS.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X82825; CAAS6032.1;
 CC EMBL: X82824; CAAS6031.1;
 CC EMBL: Z97339; CAB10369.1;
 CC EMBL: AL161542; CAB78632.1;
 CC InterPro: IPR001680; WD40.
 CC Pfam: PF00400; WD40_7.
 CC PRINTS: PR00320; GPROTEINRPT.
 CC SMART: SM00320; WD40_7.
 CC PROSITE: PS00678; WD_REPEATS_1; 2.
 CC PROSITE: PS00682; WD_REPEATS_2; 5.

DR PROSITE: PS00294; WD_REPEATS_REGION: 1.
 KW Repeat: 174
 FT REPEAT 216
 FT REPEAT 258
 FT REPEAT 300
 FT REPEAT 342
 FT REPEAT 384
 FT REPEAT 433
 FT REPEAT 486 AA; 54009 MW; 6040D3E6FCDA8A998 CRC64;
 SQ SEQUENCE

Query Match 8.2%; Score 209; DB 1; Length 486;
 Best Local Similarity 21.0%; Pred. No. 5.1e-09;
 Matches 103; Conservative 78; Mismatches 197; Gaps 17;

13 STVRNDSPPPEPESRLRHSRM-----INSNHTSPKRTIYSPDRTPPSASKFA 63
 32 SYVHGQFPPEPPEAKQILSKMKVARGVEPVSPPPQPR--INQOPSNALSLAA 89
 64 LRDINPFGRRDSSAYTLRLTALFPGDPAVPPE-----KTDSPMTLPNNRTY 118
 90 PEGSKTGKATESA-----IVGPTLRPIPKGLNTSGSKSTIIPANVSY 140
 119 KTRQSMHSLSPFMDDFVGVNHPVAKRKYPSRYVDAPALODFYLTIVDMS 178
 141 -----QKNTSTALAM-----RIPSRDPRPW--NAP-----WRN 168
 179 HNVAVLGLG-----NCYIMNCSKVTYKLCGLGYDDCVGVGNA 218
 169 YRVIQHLGVWRSVAFDPSNEMFCTGSADRTIKIDVATG-VLKLTLTGHEVGLANS 227
 219 ORGHVAVGTNGKGVQIMDAARCKIRSEGRRLVYGLAMSSP--LSSGRDKNIYR 276
 228 NHTYMFAGDQKQVCKMDEONKVRISYHGLSGVYCLALHPTLDVLLTGGRSVGRW 287
 277 DTRQEDFVSKLSGKSEVCGIKMSYDNRELAGSGDNKLFVWNOHNSQPVLYKCEHTAA 336
 288 DIRTQKQ--IFALSGDNVCSVFTFRPDPOVYVSHDTIKITWDLRYGKTMSTLTHHKS 346
 337 VKAIASPHLHGLASGGGTADRCIRFNTTNSHLSMDTGSQVCNLVWKNVE-LYS 395
 347 VRMTLHPREN--AFASASADNTKPF--SLPKGECHMLSOOKTIIINAAVEDGM 400
 396 THGSONQIIVMVPMSKLTATLTGTYRVLIALSP-----DQOTIYV 439
 401 VTGGDNCSTIFWPK-----SGHSFOSEETIYQSLSEAGIYAAQYDVTGRLVT 452
 440 GAGDETLRFW 449
 453 CEADRTIKW 462

RESULT 13
 CIAL_HUMAN STANDARD; PRT; 339 AA.
 ID CIAL_HUMAN
 AC 076071;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE WD40-repeat containing protein Cial 1.
 GN CIAL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 ON [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98225157; PubMed=9556563;
 RA Johnstone R.W., Wang J., Tommerup N., Vissing H., Roberts T., Shi Y.,
 RT "Ciao 1 is a novel WD40 protein that interacts with the tumor
 RT suppressor protein p16."
 RT J. Biol. Chem. 273:10880-10887(1998).

```

RN [2]
RA SEQUENCE FROM N.A.
RA Adams M.D., Loftis B.J., Zhou L., Crosby M., Fuhrmann J., Mason T.M.,
RA Brandon R., Kim U.J., Kerlagage A.R., Venter J.C.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RP CHARACTERIZATION.
RX MEDLINE-99367400: PubMed-10438340;
RX Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
RT "Structural organization, tissue expression, and chromosomal
RT localization of C1ao 1, a functional modulator of the Wilms' tumor
RT suppressor, WT1."
RL Immunogenetics 49:900-905(1999).
CC -1- FUNCTION: SEEMS TO SPECIFICALLY MODULATE THE TRANSACTIVATION
CC ACTIVITY OF WT1.
CC -1- SUBUNIT: INTERACTS WITH WT1.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: "C1AO" MEANS BRIDGE IN CHINESE.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC -----
DR EMBL; U6810; AAC24948.1;
DR EMBL; AC004020; AAC23493.1;
DR MIM; 604333;
DR InterPro: JPRO01680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Nuclear protein.
FT REPEAT 14 53
FT REPEAT 59 98
FT REPEAT 103 142
FT REPEAT 148 187
FT REPEAT 192 231
FT REPEAT 250 289
FT REPEAT 301 338
SQ SEQUENCE 339 AA; 37840 MW; 63A80B257A204FC8 CRC64;

Query Match
Best Local Similarity 8.1%; Score 206.5; DB 1; Length 339.
Matches 66; Conservative 39; Mismatches 83; Indels 65; Gaps 11;

QY 215 VGNAGQCHILAVGTNGKQYQIN-----DAARCKRKIRSEGRRLVGLAWS--SSILSSCG 268
DB 22 LANNPAGTILASCAGGDRIRIRIWTEGSMWICKSVLS-EGHQTIVKVAWVSPCGNTLASAS 80
QY 269 RDKN-IYQDRIRQDEP--VSKLSGRKSEVCGSLKSYDNRFELASGDNKDLFW----- 319
DB 81 FLATTCIRKKN--QDDECTTLEGEHENEYKSVAMAPSGNLTAFCSGSDKSVWWEYDEE 137
QY 320 NQHSIQPLKLCETPAAVKAIAMSPHILHGILASGGGTADRCIRFWNNTTNSHLSGMDTGS 379
DB 138 DEYEVSVALN-SHTQDYKHWVHP-SQELLAS--ASTYDPTVKLIRREEDDWVCC----- 187
QY 380 QVNCILWKSNNVNEIYSHGYSNQIITWRYPTMSKLTATLGHTRVLYLAIISPDDGIYV 439
DB 188 -----
QY 440 GAGDEPLRENNVF 452
DB 212 CSDDRIVKTRMQ 224

```

	RESULT	4	
ID	PRO4_YEAST		
AC	P20053.	STANDARD:	PTI; 465 AA.
DT	01-FEB-1991 (Rel. 17, Created)		
DT	01-FEB-1991 (Rel. 17, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, last annotation update)		
GN	U4/U6 small nuclear ribonucleoprotein PRP4		
GN	PRP4 OR RNA4 OR YPRJ78R OR p9705.6.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina;		
CC	Saccharomycetes; Saccharomycetaceae; Saccharomyces.		
OX	NBLTaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89384596; PubMed=2528686;		
RA	Petersen-Bjorn S., Soltyk A., Beggs J.D., Friesen J.D.;		
RT	"PR4 (RNA4) from Saccharomyces cerevisiae: its gene product is		
RL	associated with the U4/U6 small nuclear ribonucleoprotein particle."		
RM	Mol. Cell. Biol. 9:3698-3709(1989).		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89384597; PubMed=2528687;		
RA	Banroques J., Abelson J.N.;		
RT	"PR4: A protein of the yeast U4/U6 small nuclear ribonucleoprotein		
RL	particle.";		
RM	Mol. Cell. Biol. 9:3710-3719(1989).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288C / AB972;		
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,		
RA	Favetto A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,		
RA	Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,		
RA	Johnston L., Langston Y., Latreille P., Le T., Mardis E., Meneses S.,		
RA	Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,		
RA	Talbot A., Trevaakis E., Vignati D., Wilcox L., Wohldman P., Vaudin M.,		
Wilson R., Westerson R.;			
Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.			
-1- FUNCTION: INVOLVED IN RNA SPLICING. IS REQUIRED FOR THE			
ASSOCIATION OF U4/U6 snRNP WITH U5 snRNP IN AN EARLY STEP OF			
SPLICOSOME ASSEMBLY.			
-1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation,		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as their content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@lsb-sib.ch).		
DR	EMBL; M65957; AAA79332.1;		
DR	EMBL; M28518; AAA79011.1;		
DR	EMBL; U25842; AAB68111.1;		
DR	PIR; A32569; A32569.		
DR	GSD; S0006382; PRP4.		
DR	InterPro; IPR003648; SFM.		
DR	InterPro; IPR001680; WDA0.		
DR	pfam; PF00400; WDA0: 7.		
DR	PRINTS; PRO0320; GPOTETINBRPT.		
DR	SMART; SMO0500; SPW; 1.		
DR	SMART; SMO0320; WDA0: 7.		
DR	PROSITE; PS00678; WD_REPEATS_1; 3.		
DR	PROSITE; PS50082; WD_REPEATS_2; 4.		
DR	PROSITE; PS50294; WD_REPEATS_REGION; 1.		
KW	mRNA processing; Repeat; WD repeat.		
FT	REPEAT 216 .. 247		
FT	REPEAT 263 .. 293		
FT	REPEAT 305 .. 335		
FT	REPEAT 347 .. 377		
FT	REPEAT 391 .. 423		
FT	REPEAT 435 .. 465		
TM	WD 6.		

Wed May 22 11:09:50 2002

us-09-701-572-2.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2002, 11:46:22 ; Search time 31.11 Seconds
(without alignments)
2641.356 Million cell updates/sec

Title: US-09-701-572-2
Perfect score: 2555
Sequence: 1 MDGTGNRNPPTSTVRDNPSP.....KQNTFSEIGALSLGRTIR 475

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_prodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2555	100.0	475	10	Q9XED5 medicago va
2	2547	99.7	475	10	Q9W712 arabidopsis
3	1839	72.0	444	10	Q82740 arabidopsis
4	1825	71.4	482	10	Q91060 arabidopsis
5	1448	56.7	472	10	Q9PFY8 arabidopsis
6	1294	50.6	493	4	Q75869 homo sapien
7	1294	50.6	504	4	Q9UJH8 homo sapien
8	1286.5	50.4	496	4	Q9UM11 homo sapien
9	1286	50.3	493	13	Q42585 Q91K5 mus musculu
10	1285	50.3	493	13	Q42585 Q91K5 mus musculu
11	1276	49.9	478	5	Q18402 O18402 drosophila
12	1276	49.9	478	5	Q9W4H9 O9W4H9 drosophila
13	1274	49.9	493	4	Q9UJH8 O9UJH8 homo sapien
14	1170.5	45.8	528	10	Q94DF9 O94DF9 oryza sativ
15	1096	42.9	702	5	Q09649 Q09649 caenorhabdi
16	1066	41.7	556	3	O13286 O13286 schizosacch

17	1021.5	40.0	404	4	Q96NM8 O96NM8 homo sapien
18	1020.5	39.9	404	4	Q9UM10 Q9UM10 homo sapien
19	995	38.9	451	5	Q9W1F6 Q9W1F6 drosophila
20	985	38.6	421	3	Q94423 Q94423 schizosacch
21	935.5	36.6	457	10	Q82669 Q82669 brassica na
22	923	36.1	447	10	Q9S718 Q9S718 arabidopsis
23	920.5	36.0	457	10	Q9S718 Q9S718 arabidopsis
24	920.5	36.0	460	10	Q9S50K2 Q9S50K2 arabidopsis
25	917	35.9	526	5	Q24044 Q24044 drosophila
26	904	35.4	450	10	Q70380 Q70380 rattus norv
27	901.5	35.2	499	11	Q23919 Q23919 rattus norv
28	900	35.2	499	11	Q62623 Q62623 rattus norv
29	895	35.0	499	11	Q991K3 Q991K3 mus musculu
30	894.5	35.0	499	4	Q9W4H9 Q9W4H9 drosophila
31	893.5	35.0	499	4	Q12834 Q12834 homo sapien
32	892.5	34.9	499	11	Q9UJH8 Q9UJH8 homo sapien
33	891	34.9	499	11	Q9UJH8 Q9UJH8 mus musculu
34	880.5	34.5	442	10	Q9S7H3 Q9S7H3 arabidopsis
35	876	34.3	440	10	Q04634 Q04634 arabidopsis
36	868.5	34.0	522	5	O61588 O61588 spissula sol
37	860.5	33.7	507	13	Q93289 Q93289 xenopus lae
38	851	33.3	356	11	Q9CT59 Q9CT59 mus musculu
39	782	30.6	838	5	Q9N748 Q9N748 leishmania
40	777.5	30.4	377	4	Q9Y2T8 Q9Y2T8 homo sapien
41	777.5	30.4	488	3	P78972 P78972 schizosacch
42	666	26.1	310	6	Q9BDJ9 Q9BDJ9 sus scrofa
43	565	22.1	142	4	Q9UJH8 Q9UJH8 homo sapien
44	555.5	21.7	424	5	Q00809 Q00809 tritichomo
45	458.5	17.9	593	3	Q96VQ3 Q96VQ3 saccharomyc

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	475 AA.
Q9XED5	Q9XED5			
AC	Q9XED5			
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	CELL CYCLE SWITCH PROTEIN.			
GN	CCS52.			
OS	Medicago varia.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
CC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.			
OX	NCBI_TaxID=36902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=NODE;			
RA	Cebolla A., Kondorosi E., Kondorosi E., Kondorosi E.;			
RT	"A plant cell cycle regulator that controls endoreduplication and cell			
RT	differentiation."			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).			
DR	EMBL: AF079404; AAD22612.1;			
DR	InterPro: IPR000002; F1ZYZ.			
DR	InterPro: IPR001680; WD40.			
DR	Pfam: PF00400; WD40_5.			
DR	PRINTS: PR00320; GPROTEINRPT.			
DR	PRODOM: PD004563; F1ZYZ; 1.			
DR	SMART: SM00320; WD40_5.			
DR	PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.			
DR	PROSITE: PS00082; WD_REPEATS_2; 3.			
DR	PROSITE: PS0294; WD_REPEATS_REGION; 1.			
KW	REPEAT; WD repeat.			
SQ	SEQUENCE 475 AA; 52542 MW; 8E54F0DAFAC018FB CRC64;			

Query Match 100.0%; Score 2555; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.2e-211;
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMSINSNHYTSPRTIYSDRFIFRSAS 60
 DB 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMSINSNHYTSPRTIYSDRFIFRSAS 60
 QY 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPVPEKTDSPKMTLPNRIFFRYKT 120
 DB 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPVPEKTDSPKMTLPNRIFFRYKT 120
 QY 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 DB 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 QY 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHILAVGNNCKVOIWDAR 240
 DB 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHILAVGNNCKVOIWDAR 240
 QY 241 CKIRSMEGHRLRVGALWASSSLSSGGDKNITVORDIRTOEDFVSKLSGHSKSEVGLKW 300
 DB 241 CKIRSMEGHRLRVGALWASSSLSSGGDKNITVORDIRTOEDFVSKLSGHSKSEVGLKW 300
 QY 301 SYDNRELASGGDNKLFVWNOHSTOPLVKCEHTAAVKAIAKMSPHLGLLASSGGTADRC 360
 DB 301 SYDNRELASGGDNKLFVWNOHSTOPLVKCEHTAAVKAIAKMSPHLGLLASSGGTADRC 360
 QY 361 IREFWNTTNSHLSGMDTSGOVCNLYWSKNNVELVSTHGYSONOIIYWRYPYMSKLTATL 420
 DB 361 IREFWNTTNSHLSGMDTSGOVCNLYWSKNNVELVSTHGYSONOIIYWRYPYMSKLTATL 420
 QY 421 HTYRVLYLAISPDDGQITVAGDETLRFVWVFPSPKSONTESIGLSIGRTTIR 475
 DB 421 HTYRVLYLAISPDDGQITVAGDETLRFVWVFPSPKSONTESIGLSIGRTTIR 475

RESULT 2
 Q9M712 PRELIMINARY: PRT: 475 AA.
 AC Q9M712;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE WD-REPEAT CELL CYCLE REGULATORY PROTEIN.
 GN CCS52.
 OS Medicago truncatula (Barrel medic).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
 RN NCBL_TaxID=3880;
 RC SEQUENCE FROM N.A.
 RA STRAIN=GHOR;
 RA Cebolla A., Vinardell J.M., Olah B., Kiss E., Roudier F.,
 RA Kondorosi A., Kondorosi E.;
 RT "The mitotic inhibitor ccs52 is required for endoreplication and
 RT polyploid-dependent cell enlargement in plants.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF134835; AAF37386.1;
 DR InterPro: IPR000002; F1zzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.6.
 DR PRINTS: PRO0320; GPROTEINBRPT.
 DR ProDom: PD004563; F1zzy.1.
 DR SMART: SM00320; WD40.5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 475 AA: 52457 MW: 2778EB31FD20D5C2 CRC64;

Query Match 99.7%; Score 2547; DB 10; Length 475;
 Best Local Similarity 99.6%; Pred. No. 5.8e-211;

Matches 473; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMSINSNHYTSPRTIYSDRFIFRSAS 60
 DB 1 MDGTGNRPPTSTYRONSPPPEPSPESLRHSVMSINSNHYTSPRTIYSDRFIFRSAS 60
 QY 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPVPEKTDSPKMTLPNRIFFRYKT 120
 DB 61 KPALFDINTPTGEGDDSSAYTTLLRTALFGPDVAGPVPEKTDSPKMTLPNRIFFRYKT 120
 QY 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 DB 121 ETRQSMHSLSPFMDDDVPVGNHSPKAPRKPRSPYKVLADAPALODDFYLNLDVMSHN 180
 QY 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHILAVGNNCKVOIWDAR 240
 DB 181 VLAAGLNCYVLLWNAACSSKVTKLCDLGVDDCVCSYGMORGHILAVGNNCKVOIWDAR 240
 QY 241 CKIRSMEGHRLRVGALWASSSLSSGGDKNITVORDIRTOEDFVSKLSGHSKSEVGLKW 300
 DB 241 CKIRSMEGHRLRVGALWASSSLSSGGDKNITVORDIRTOEDFVSKLSGHSKSEVGLKW 300
 QY 301 SYDNRELASGGDNKLFVWNOHSTOPLVKCEHTAAVKAIAKMSPHLGLLASSGGTADRC 360
 DB 301 SYDNRELASGGDNKLFVWNOHSTOPLVKCEHTAAVKAIAKMSPHLGLLASSGGTADRC 360
 QY 361 IREFWNTTNSHLSGMDTSGOVCNLYWSKNNVELVSTHGYSONOIIYWRYPYMSKLTATL 420
 DB 361 IREFWNTTNSHLSGMDTSGOVCNLYWSKNNVELVSTHGYSONOIIYWRYPYMSKLTATL 420
 QY 421 HTYRVLYLAISPDDGQITVAGDETLRFVWVFPSPKSONTESIGLSIGRTTIR 475
 DB 421 HTYRVLYLAISPDDGQITVAGDETLRFVWVFPSPKSONTESIGLSIGRTTIR 475

RESULT 3
 O82740 PRELIMINARY: PRT: 444 AA.
 AC O82740;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE PUTATIVE F1ZZY-RELATED PROTEIN.
 GN F7H19.90 OR ATAG22910.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBL_TaxID=3702;
 RC SEQUENCE FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Scheller C., Bevan M.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AL031018; CA119806.1;
 DR EMBL: AL161558; CAB79246.1;
 DR InterPro: IPR000002; F1zzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.7.
 DR PRINTS: PRO0320; GPROTEINBRPT.
 DR ProDom: PD004563; F1zzy.1.
 DR SMART: SM00320; WD40.5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.

Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
1 MDGTGNRPPTSTVRDNDSP-----PEPSPELRVSRMTSNHNTSPSKRTYSIDRFPIS	71.4%;	72.9%;	1825;	DB 10;	482;	30;	7
1 MEDESTTEPKKKSQNLPPSMNRPTVSLES--RINLIDSNHYSPPSKPYISDRFPIS	46;	54;					
57 RSASKRALFDI--NTPP--EGRDDSSAYTTLLRLALFGPOVAGPVPTEKTDSPMTLPN							
59 RSGSNALPELDLSSPPKMKKCKEDGAGSYASLLKTLAF-----GVTPKSDVYNGCFSP							
113 RNIFKRYKTEKROSMHSLSPENDDDFFVGVMHSPVKAPKRPVPSPPKVLADALADDEFLN							
114 GNIFREFTEFORSNLNTPPF--DSVDVSGVSPSPKSPKRLKSLFKVLDAPALODEFLN							
173 LVDMSNNVLAVLGNVCYLLANACSK-----YTKLGLGNDCCVSGMA							
173 LVDMSAQVAVLAVGAGNVCYLLANACSKLYSPFMETSCYGLVYTKLGLGDEFVSGMA							
219 ORGTHLANGTNNGKVOJIMDARCKTRISMEGHRLVAGALAMSSLSLSSGGKGNITQRI							
233 LRGTHLALTSISGVIQWLDVLRCKNIRTMGKRLVAGALAMSSVLSLSSGRDSTLQRI							
279 RTQEDFVSKLSGKSEVCGELKMSYDNRETLASGNDNKLFFVMNQHSTQPVLFCEHAAYK							
293 RTQEDHVSRLKGGKSEICGLKMSSDNRETLASGNDNKLFFVMNQHSTQPVLFCEHAAYK							
339 AIAWSPHLHGLLASGGGTADRCIRFMWNTTNSHLSGNDTGSOVCLVMSKNVNELVSTHG							
353 AIAWSPHFLILASGGGTADRCIRFMWNTTNTNHLNCVDTISOVCLVMSKNVNELVSTHG							
399 YSONDIYWRPMTSKLATTIGTATRVLYLAIISPDGQITVITGAGDETLFFWVVPSPSKQ							
413 YSONDIYWKYPTMSKLATVLTGHSTVNLVLANSPDGQITVITGAGDETLFRMWNFSPSKQ							

RA Sato S., Kotani H., Nakamura Y., Kaneke T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabeta S.,
 RT Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT clones.
 CC DNA Res. 4:215-230(1997).
 DR -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AB005230. Babi112.1.
 DR InterPro: IPR000002; F14zy.
 DR Pfam: PR001680; WD40.
 DR ProDom: PD004563; F14zy; 7.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 4.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat
 SQ SEQUENCE 472 AA; 51330 MW; A9C6BAD79B5D4FB3 CRC64;

Query Match
 Best Local Similarity 56.7%; Score 1448; DB 10; Length 472;
 Matches 277; Conservative 60; Mismatches 90; Indels 22; Gaps 7;

OY 23 EPSESLRHVSRMINSNHTSPRTIYSDRIPESRSKPFALDI-----NTEGRDSDS 77
 DB 25 ETSSSFRCISL-----SSPKSCSDREIFCRSSRLHAFDLOKEPTTV--KEGG 76
 OY 78 SNAVTLRLTALFEGDVAGP-VTEE--KTDSPTMLPNINIRYTTETROSMSLSPMD 134
 DB 77 NAYSRLKLSLFGSDRASPPLSPAGGQSASPSMPCITMLRKTRETSNPS-SPFS 135
 OY 135 DDFVGVNHS-----PVKARVRSRPRVLAAPLADDFYLVDMSSNNVLAVALGNC 189
 DB 136 SILGNDGSHSDSPPPKPRKPKRKHVLDAPLDDDFYLVDMSSNNVLAVALGNC 189
 OY 190 VYLMAACSSKYTKLGDVDCVCSVAGSGTHLAVGTNNGVQIWDARCKTIRSMEG 249
 DB 196 VYLTMASSNKVTKLGDVDCVCSVAGSGTHLAVGTNNGVQIWDARCKTIRSMEG 249
 OY 250 HRLVALGALSSSLSSGGRKNITORDIRTOEDFVSKISGKSEYCGILKMSYDNRELAS 309
 DB 256 HQRTGVLAAMNRILSSGSRDRNLQDIDVQSDPYLVKLGKSVCCILKMSHDELAS 315
 OY 310 GGNDKLFVWNOHSTOPLVLCYCEHTAAVKAIAWSPHGLLASGGVADRIRFWNTTN 369
 DB 316 GGNDKLFVWNOHSTOPLVLCYCEHTAAVKAIAWSPHGLLASGGVADRIRFWNTTN 369
 OY 370 SHLSMDTGSQVCLVWKNVNLSTHGYSONOIIWRYPTMSKATLGLHYRVLXLA 429
 DB 376 NQLSIDTGSQVCLVWKNVNLSTHGYSONOIIWRYPTMSKATLGLHYRVLXLA 429
 OY 430 ISPDGQIYVAGDETLRWNVFSPKSO 458
 DB 436 TSPDQGIYVAGDETLRWNVFSPKSO 458

RESULT 6
 ID 075869 PRELIMINARY; PRT: 493 AA.
 AC 075869.
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE R33374_1 (HYPOTHEICAL 54.8 KDA PROTEIN).
 GN F14.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP NCBI_Taxid-9606;
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilhagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Danganan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Actix C., Lucase T., Frankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas R., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arelano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
 RT serine protease gene cluster."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Jin D.-Y., Jeang K.-T.;
 RT "Characterization of human fizzy-related protein."
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP TISSUE-BRAIN, NEUROBLASTOMA;
 RA Strausberg R.;
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AC005787; AAC62835.1;
 DR EMBL: AF083810; AAC62836.1;
 DR EMBL: BC013413; AAC52030.1;
 DR InterPro: IPR000002; F14zy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PR001680; WD40.
 DR PRINTS: PR00320; GPROTEINRPT.
 DR ProDom: PD004563; F14zy; 7.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Hypothetical protein.
 SQ SEQUENCE 493 AA; 54765 MW; 766A947EB8BA6D5C CRC64;

Query Match
 Best Local Similarity 50.6%; Score 1294; DB 4; Length 493;
 Matches 269; Conservative 66; Mismatches 106; Indels 54; Gaps 14;

OY 6 NRPSTSVDRNSPPPESESLRHVSRMINSNHTSPRTIYSDRIPESRSKPFAL- 64
 DB 17 NENTPRVTEMRKRLTPASSPV-SPSK--HGDRIPSPRAGAMSVN 60
 OY 65 -PDIN-----TEBGR--DDSSS-----AYTLRLTALFEG--PDVAGVTPREKTDSPS 107
 DB 61 FHRINENKSPSONRRAKDATSONGKDLATSLKKELLGAGIEKVDOPQEDRLDPS 120
 OY 108 WTLNRRNIFRYKTETRQSK-----HSLSPMDDFVGVNHSPPVAKPRVRSYK 158
 DB 121 -TEPKGLFTYSLSTRKSSPDGNDVSPYSLSPVSNKS--QKLSRPRKPRKRSKIPFK 177
 OY 159 VLDAPALDDDFYLVDMSSNNVLAVALGNCVYLVMAACSSKYTKLGDV--DCVCSGVN 217
 DB 178 VLDAPALDDDFYLVDMSSNNVLAVALGNCVYLVMAACSSKYTKLGDV--DCVCSGVN 217
 OY 218 AORGTHLAVGNNKQVQIMDAARCKIRSMEGHRLVAGALMASSSLSSGGRKNITOND 277
 DB 238 SERGLVAVGTHKCFVQIMDAAGKRLSMLEBCHRAVALAMNAEQLSSGRMRLDND 297
 OY 278 IRT---OEDEFVSKLSGKSEYCGILKMSYDNRELASGGNDKLFVWNOHSTOPLVLCYCEHT 334
 DB 298 IRTPELQSE--RLDQHQEVGGLKMSWDHQLLASGGNDKLLVWNHSSLSVQDYTEHL 355
 OY 335 AAVKALAMSPHGLLASGGVADRIRFWNTTNHLSMDTGSQVCLVWKNVNLSTHGYSONOIIWRYPTMSKATLGLHYRVLXLA 394
 DB 356 AAVKALAMSPHGLLASGGVADRIRFWNTTNHLSMDTGSQVCLVWKNVNLSTHGYSONOIIWRYPTMSKATLGLHYRVLXLA 415
 OY 395 STHGYSONOIIWRYPTMSKATLGLHYRVLXLAISPDGQIYVAGDETLRWNVFSPKSO 454
 DB 416 STHGYSONOIIWRYPTMSKATLGLHYRVLXLAISPDGQIYVAGDETLRWNVFSPKSO 473


```

Db 356 AAVKAIAMSPHOGHLLASGGCTADRCIRFWMNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
QY 395 STHGYSQNOIYWRPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFNVPFS 454
Db 416 STHGYSQNOIYWRPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFNVPFS 454
QY 455 PKSONTESE-IGALSL 469
Db 476 TRSTKYKMSVSVLNL 491

RESULT 9
ID 09RKS PRELIMINARY; PRT: 493 AA.
AC 09RKS:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE FIZZY-RELATED PROTEIN (FZRL PROTEIN).
GN FZRL OR FYR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_Taxid=10090.
RP SEQUENCE FROM N.A.
RA Jin D.-Y., Jeang K.-T.
RT "Characterization of mouse fizzy-related protein."
RN Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RP [2]
RA Strauberg R.
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF083809; AAD52029.1;
DR MGD: BC006616; AAH06616.1;
DR InterPro: IPR000002; Fizzy.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR ProDom: PD004563; Fizzy; 1.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 3.
DR Repeat: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SO SEQUENCE 493 AA; 54688 MM; B5DC6653D74D9A5 CRC64;

```

Query Match 50.3%; Score 1286; DB 11; Length 493;
 Best Local Similarity 53.9%; Pled. No. 2,5e-102;
 Matches 267; Conservative 69; Mismatches 105; Indels 54; Gaps 14;

```

Db 6 NNNPPTSTVDNSPPPEPSRLRVSRMINSNHYTSPRTIYSDRFPISRSKAFAL- 64
QY 17 NENIVPCVSEMRRTLTPTANSVPS-----SPSK--HGDRFIPSRAGANMSVN 60
QY 65 -FDIN---TPTEGR--DDSS-----AYTLRLTALFG---PDVAGPYTPKTDSPS 107
Db 61 FHRINENKSPSONRKAKDATSDNGKGLAYSAALKNELLGAGIEKVOPOEDRRLPS 107
QY 108 MTLPRNIRFRYTKETRO-----HSLSPFMDDFVPGVNSPVKAPRKVPSPYK 120
Db 121 -TPEHGLFTYSLSSKRSSPDGNDVSPYSLSPSNKS--OKLSPRKPRKTKIPFK 177
QY 159 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCYVLMAGSKYTKICLDGYD-DCVGSVGM 217
Db 178 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCYVLMAGSKYTKICLDGYD-DCVGSVGM 217
QY 218 AORGHILAVGNGNGKQVIMDAACKIRKMBEHRIRVAGALAMSSSLSSGGRKNITQOD 277
Db 238 SERGNLVAETHNGFVOIMDAAGKLMLESGHARVAGALAMADSLSSGGRKNITQOD 297

```

```

QY 278 IRT---QEDFVSKLGHKSEVCGKMSYDNRELASGNDKLFVWNGHSTOPLKYCEHT 334
Db 298 IRTPELOSE--RLQGRDEVCGLKMSWDHOLASGNDKMLVWNNSSLSPTVOQYTEHL 355
QY 335 AAVKAIAMSPHOGHLLASGGCTADRCIRFWMNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
Db 356 AAVKAIAMSPHOGHLLASGGCTADRCIRFWMNTLTCGLOPIDTSSOVCLAMSKHANLEY 415
QY 395 STHGYSQNOIYWRPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFNVPFS 454
Db 416 STHGYSQNOIYWRPTMSKLTATLGHTRVYLALSPDGGOTITGAGDETLRFNVPFS 454
QY 455 PKSONTESE-IGALSL 469
Db 474 SKTRSTKESVSVLNL 488

```

```

RESULT 10
ID 042585 PRELIMINARY; PRT: 493 AA.
AC 042585:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
GN FIZZY-RELATED PROTEIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
RN NCBI_Taxid=8355;
RP SEQUENCE FROM N.A.
RA TISSUE-OVARY.
RC MEDLINE-97433078; PubMed-9288747;
RA Sigrist S.J., Lehner C.F.
RT "Xenopus laevis fizzy-related down-regulates mitotic cyclins and is
RT required for cell proliferation arrest and entry into endocycles."
RL Cell 90:671-681(1997).
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: Y14163; CA474576.1;
DR InterPro: IPR000002; Fizzy.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR ProDom: PD004563; GPROTEINRPT.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 3.
DR Repeat: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SO SEQUENCE 493 AA; 54699 MM; BA19741AD3CB6A7 CRC64;

```

Query Match 50.3%; Score 1285; DB 13; Length 493;
 Best Local Similarity 53.7%; Pled. No. 3e-102;
 Matches 266; Conservative 70; Mismatches 105; Indels 54; Gaps 14;

```

Db 6 NNNPPTSTVDNSPPPEPSRLRVSRMINSNHYTSPRTIYSDRFPISRSKAFAL- 64
QY 17 NENIVPCVSEMRRTLTPTANSVPS-----SPSK--HGDRFIPSRAGANMSVN 60
QY 65 -FDIN---TPTEGR--DDSS-----AYTLRLTALFG---PDVAGPYTPKTDSPS 107
Db 61 FHRINENKSPSONRKAKDATSDNGKGLAYSAALKNELLGAGIEKVOPOEDRRLPS 107
QY 108 MTLPRNIRFRYTKETRO-----HSLSPFMDDFVPGVNSPVKAPRKVPSPYK 120
Db 121 -TPEHGLFTYSLSSKRSSPDGNDVSPYSLSPSNKS--OKLSPRKPRKTKIPFK 177
QY 159 VLDAPALODDFYLNLYDMSSHNLYAVGLGNCYVLMAGSKYTKICLDGYD-DCVGSVGM 217

```


RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Smith H.O.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AEO03431; AAF45973.1; -
 DR Flybase: FBgn0003200; rap.
 DR InterPro: IPR000002; F1zzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.
 DR ProDom: PD004563; F1zzy; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; 2.
 DR PROSITE: PS0082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 478 AA; 53507 MW; 0E6098740897CF76 CRC64;

Query Match 49.98; Score 1276; DB 5; Length 478;
 Best Local Similarity 54.98; Pred. No. 1.7e-101;
 Matches 260; Conservative 62; Mismatches 114; Indels 38; Gaps 11;

DB 22 PEPSPELRVSRKINS--NHVTSPTRTYSDRTFSPSS--ASKRALFDIN-----68
 DB 4 PEYKRLKHYSPVARNLNFNFESSITPTSLDRTPICAYNNMOTJNA--SINKSDNSP 61
 QY 69 -TPTEGRD-----DSSAVTLLRTALFGRPV-----AGVTEPKTDSPTLPRNITR 117
 DB 62 QTSKRDGCGEATROSLAYSCILNKGSAIDVKTAGEENENAYTRA--AKSLK 118
 QY 118 YKTEQNSM-----HSLSPFMDDFVPGVNSPVKAPRKVPSPYVLDAPLDDFYLN 172
 DB 119 YQSPRTQDYNGCEPYSLSPPVSAKS--OKLLRSPRKATRKISRPFLVDLAPLDDFYLN 176
 QY 173 LVDMSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSVMAQRGTHLAVGTNG 231
 DB 177 LVDMSQNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSVMAQRGTHLAVGTNG 231
 QY 232 KVALIDARCKKIRSMGCHRLRGALAMSSSLSSGGRDKNTYORDT--OEDVSVKLSG 290
 DB 237 YVTVMDVAANKQIKKLGHSARVALAMNSDILSSGSRDMILQDRTTPLOSERLAG 296
 QY 291 HKSPVCLAKSYNRELASGNDNKLFWNNHSTOPTVLYKCEHRAVKAIAVSPHLLGL 350
 DB 297 HRDVEGCLKMSPDNOYLASGNDNRLVWNGHSAVPOSTTEHMAAKAIAVSPHLLGL 356
 QY 351 ASGGGTADRCIRFWNTTNSHLSGMDTSGOVCLVMSKNVNLVSTHGSQNOIIVMYRP 410
 DB 357 ASGGGTADRCIRFWNTTNSHLSGMDTSGOVCLVMSKNVNLVSTHGSQNOIIVMYRP 410
 QY 411 TMSKLTATLGHVRYLYLAISPQDQITVAGDETLFNNVFPSPSQTSEET 464
 DB 417 SLTVAKLTGHSYRVLTLALSPDEALVTGAGDETLRFNNVFPSPSQSKSKENKSV 470

RESULT 13
 ID 090196 PRELIMINARY; PRT; 493 AA.
 AC 090196;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 01-DEC-2001 (Tremblrel. 13, Last annotation update)
 GN F1Zzy-RELATED PROTEIN HOMOLOG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Sudo T., Sawa H.;
 RT "Identification of a human homolog of the *Drosophila* fuzzy-related
 protein."
 RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF080397; AAF20266.1; -
 DR InterPro: IPR000002; F1zzy.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.
 DR ProDom: PD004563; F1zzy; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN.2.
 DR PROSITE: PS0082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 493 AA; 54905 MW; 9DE5DA138C19383C CRC64;

Query Match 49.98; Score 1274; DB 4; Length 493;
 Best Local Similarity 53.78; Pred. No. 2.7e-101;
 Matches 266; Conservative 67; Mismatches 108; Indels 54; Gaps 14;

QY 6 NRPPTSTVTRDNPSPSPSPSLRVSRKINSNHVTSPTRTYSDRTFSPSSASKRAL- 64
 DB 17 NENTMRYTEMKRTILTPASSPV-----SPSK--HGDRFTSRAGAMSVN 60
 QY 65 -FDIN-----TPTEGR--DSSS-----AVTLLRTALFG--PDVAGPVTPKTSOPS 107
 DB 61 FRIENEMSPSONKAKADATSDNGKGLAYSLNELLGAGIERKVDQPTEDRRLOPS 120
 QY 108 MTLPRNIRFKYKTEKROSM-----HSLSPFMDDFVPGVNSPVKAPRKVPSPYK 158
 DB 121 -TEPKGLFTYSLSTKSSPDCCNAXVPSLSPPVSAKS--OKLLRSPRKATRKISRPFL 177
 QY 159 VLDPALDDFYLYNLDVMSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSVGM 217
 DB 178 VLDPALDDFYLYNLDVMSHNVLAVALGNCVYLLNACSSKVTLCGLVD--DCCVSVGM 217
 QY 218 AQKSTHLAVTNGKQVQIDARCKKIRSMGCHRLRGALAMSSSLSSGGRDKNTYORD 277
 DB 238 SERGNLVAAGTHGFPQVMDASAGKILMSLGHARVAWALANNAEOLSSGSRDMILQD 297
 QY 278 IRT--OEDVSVKLSGHSVCCGLKMSYDNRELASGNDNKLFWNNHSTOPTVLYKCEH 334
 DB 298 IRTPLQSF--RLDGHQEVCCGLKMSYDNRELASGNDNKLFWNNHSTOPTVLYKCEH 334
 QY 335 AAVKAIAMSPHLLGLASGGGTADRCIRFWNTTNSHLSGMDTSGOVCLVMSKNVNLV 394
 DB 356 AAVKAIAMSPHLLGLASGGGTADRCIRFWNTTNSHLSGMDTSGOVCLVMSKNVNLV 394
 QY 395 STHGSQNOIIVMYRPTMSKLTATLGHVRYLYLAISPQDQITVAGDETLFNNVFPSP 415
 DB 416 STHGSQNOIIVMYRPTMSKLTATLGHVRYLYLAISPQDQITVAGDETLFNNVFPSP 415
 QY 455 PKSQNTSEIGALSL 469
 DB 474 SKTRSTKESVSVNL 488

RESULT 14
 ID 094DF9 PRELIMINARY; PRT; 528 AA.
 AC 094DF9;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DE PUTATIVE CELL CYCLE SWITCH PROTEIN.
 GN P0518C01.29.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

DR EMBL: Z47358: CAA87433.1: -
DR NormBep: ZK1307.6: CR01695.
DR InterPro: IPR000002: Fizzy.
DR InterPro: IPR001680: WD40.
DR Pfam: PF00400: WD40; 6.
DR PRINTS: PR00320: GPROTEINBRPT.
DR ProDom: PD004563: Fizzy; 1.
DR SMART: SM00320: WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS00682; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Hypothetical protein.
FT DOMAIN 116 119 POLY-SER.
FT 149 154 POLY-SER.
SQ SEQUENCE 702 AA; 76403 MW; B4B78432467FB558 CRC64;

Query Match 42.9%; Score 1096; DB 5; Length 702;
Best Local Similarity 48.7%; Pred. No. 9.5e-86;
Matches 227; Conservative 59; Mismatches 134; Indels 46; Gaps 12.

QY 22 PESPES-LRHVSRMINSNHYTSPRTIYSDRFIPSRSAKFALEINTPEGRDSSSA 80
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 228 PQPTQESGTFGILIPOLERNAYGSDVTPY-----VVAEGIIIV-TISGASEPTMA 279

QY 81 YTTLTALTFGPVA---GPV-----TPEKTSPSMTL-----PKNIFPKYKET 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 QMMEPRLRGADSVADILPSPAASINGDTPPKAPPALPLSPIVOKOSPARSLFTYSAKT 339

QY 123 -----RGSMSHLSPFMDDDFVPCV-----HSFVKRPRKVPSPYKVLDAALODDFY 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 TPVAKYGQAATATTATSPF-GGPF--GVDSQRLLRTPRKPIRKVPKNPYKVLDAPELDODDFY 396

QY 171 LNIYDMSHNVLAIVGAGNCVYLIMNACSSKVTIKLDLG---VDPCVCSVGMAORGTHLAVG 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 397 LNIYDMSNQQLSTVGLAACVYLMSATFSQYIKLDLGOTNEQDQYTSVQMKDKGDLAVG 456

QY 228 TNNKQVQIMDAARCKKIRSMESGRLRVGALANSSLSGGGRKNIYQORDITQEE-DEVS 286
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 457 TSKGVTDIMDVTTQKKTRELTLGSHSRVGLAMNADTICSSGRDRTIMHRDIRDNDMDGR 516

QY 287 KLSGHSKSEVCGLKMSIDNRELASGGNDKLFVWNQHSIQPLKYCEHTAAVKALAMSPHL 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 517 KLTNHRQEVCGLKMSDPDKQLASGGNDQLLVMNLRNEPIQYTYTOHNAAVKALAMSPHL 576

QY 347 HGLLASGGGTADRCIRFNNWTTNSHLSGMDGDSQVCNLWMSKNENELVSTHGSONQIIV 406
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 577 HGLLVSGGGTADRCLEFNNWTTLTAQPPQCYVDIGSOVCNWMKSHSELVSTHGTSFNNVII 636

QY 407 WRPYTMSKIATLTGHTYRVLYLAISPDQGIIVTGADETLRFNNVF 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 637 WKRPYSIQPYTKLVGHQYRVLYLAMSPDDESIIVTGADETLRFNNVF 682

Search completed: May 21, 2002, 11:50:05
Job time: 223 sec

```

